

**A GENOME-WIDE ASSOCIATION STUDY OF SURVIVAL IN HEAD AND
NECK SQUAMOUS CELL CARCINOMA**

by
Jennifer R. Wang, BHSc., M.D.

A thesis submitted to Johns Hopkins University in conformity with the requirements for
the degree of Master of Epidemiology

Baltimore, Maryland

August, 2015

© 2015 Jennifer R. Wang
All Rights Reserved

Abstract

Head and neck squamous cell carcinoma (HNSCC) is a heterogeneous disease in terms of survival outcomes. Moreover, multimodality treatment often leads to significant morbidity. As such, there is an urgent need for personalized prognostication and treatment strategies. It has been hypothesized that germline variations may account for some of the heterogeneity observed in individual survival outcomes as they may influence tumor progression and treatment response. The objective of this study was to utilize a genome-wide approach to examine the association between germline variants and overall survival in HNSCC. Whole blood samples from 1145 patients with newly diagnosed primary HNSCC (oral cavity, oropharynx, and hypopharynx) were collected. A custom-designed Illumina Oncoarray BeadChip was used to interrogate 533,631 single nucleotide polymorphisms (SNPs) across the genome. After quality control measures, survival analysis using multivariate cox proportional hazards models was conducted on 768 individuals of European descent. No variant reached the genome-wide threshold for significance. However, two loci' associations with overall survival were suggestive of significance (p values $< 1 \times 10^{-6}$). Under the recessive model, rs1974051 located at 6p12.3 was associated with decreased survival (HR, 2.09; 95% CI, 1.57-2.79). The other two significant SNPs were in high linkage disequilibrium (rs10812227 and rs12237653) with one another and were protective variants located at 9p24.2 (HR, 0.32; 95% CI, 0.20-0.50 for each). Exploratory sensitivity analyses based on anatomical sub-site revealed two SNPs that reached genome-wide significance within the oral cavity (rs7862541) and oropharynx (rs7862541), suggesting that certain molecular pathways may be more

important in determining prognosis in one anatomical sub-site compared to another. In conclusion, this study identified novel germline variants located within loci known to be altered in HNSCC that may influence prognosis in European patients. Given their ease of detection, germline variants are ideal biomarkers for use in clinical settings. Further functional and validation studies are needed in order to determine the contribution, if any, of these candidate variants to prognostication in all or a subset of HNSCC patients.

Table of Contents

Abstract	ii
List of Tables and Figures.....	v
Preface.....	vi
1. Introduction.....	1
1.1 Epidemiology	1
1.2 Diagnosis and Treatment	2
1.3 Prognosis.....	3
2. Literature Review.....	4
2.1 Genetics of HNSCC	4
2.2 Germline Variations and HNSCC	5
2.3 Germline Variations and HNSCC Survival.....	7
3. Study Objective.....	8
4. Methods.....	9
4.1. Study Enrollment and Databases.....	9
4.2. Ethics Approval	10
4.3. Study Population	10
4.4 Data Collection	11
4.4.1. Demographic and Clinical Information.....	11
4.4.2. Blood Collection and DNA Extraction.....	11
4.5 Genotyping.....	12
4.6 Quality Control	12
4.6.1 Gender Check.....	13
4.6.2 Relatedness.....	13
4.6.3 Genotype call rates by individual and SNP	14
4.6.4 Minor Allele Frequency	14
4.6.5 Batch Effects.....	14
4.6.6 Population Substructure	14
4.6.7 Heterozygosity	15
4.6.8 Hardy-Weinberg Equilibrium	15
4.7 Statistical Analysis	15
5. Results.....	16
5.1 Quality Control	16
5.2 Patient Characteristics	18
5.3 Germline Variants Associated with HNSCC Survival.....	20
5.4 Sensitivity Analyses Based on Anatomical Sub-site.....	24
6. Discussion	25
6.1 Germline Variations and Cancer Survival	25
6.2 Germline Variants Associated with HNSCC Survival.....	26
6.3 Limitations.....	29
7. Conclusions.....	30
References.....	32
Supplementary Information	42

List of Tables

Table 1 Characteristics of patients eligible for survival analysis	24
Table 2 SNPs associated with overall survival in European HNSCC patients	42
Table 3 Significant SNPs in the oropharynx sub-site in European HNSCC patients	29
Table 4 Significant SNPs in the oral cavity sub-site in European HNSCC patients	30

List of Figures

Figure 1 Summary of germline variations associated with HNSCC	13
Figure 2 Summary of sample quality control.....	22
Figure 3 Summary of SNP quality control.....	23
Figure 4. Q-Q plot of GWAS for overall survival in European HNSCC patients.....	26
Figure 5 Manhattan plot of GWAS for overall survival in European HNSCC patients.	26
Figure 6 Unadjusted Kaplan-Meier curves for overall survival by genotype.....	27
Supplementary Figure 1 Study overview.....	47
Supplementary Figure 2 Pairwise IBD before and after removal of related pairs	48
Supplementary Figure 3 Distribution of sample call rate before and after sample-based QC	49
Supplementary Figure 4 Distribution of SNP call rate before and after marker-based QC	50
Supplementary Figure 5. Distribution of minor allele frequencies before and after SNP and sample-based QC	51
Supplementary Figure 6 First two principal components of the study population with and without reference HapMap 3 populations	52
Supplementary Figure 7 First two principal components of the study population Europeans with and without reference HapMap 3 CEU and TSI populations	53
Supplementary Figure 8 Heterozygosity rate vs. proportion of missing SNPs per sample after SNP and sample-based QC	54
Supplementary Figure 9 Minor allele frequencies and sample call rates across genotyping plates	55
Supplementary Figure 10 Regional plots for top snps across all anatomical sub-sites .	56
Supplementary Figure 11 Regional plots for the most significant SNP in oropharynx sub-site	57
Supplementary Figure 12 Regional plots for the most significant SNP in the oral cavity sub-site	57

Preface

This thesis is original, unpublished, independent work by the author, J.R.Wang.

This project was based on data generated through the OncoChip Initiative within The Genetic Association and Mechanisms in Oncology (GAME-ON) consortium and the International Head and Neck Cancer Epidemiology Consortium (INHANCE).

Genotyping was funded by the National Institute of Dental and Craniofacial Research (NIDCR) through the Center for Inherited Disease Research (CIDR): 1X01 HG.

Additional financial support for the project was provided by the American Head and Neck Society (Ballantyne Resident Research Grant) and the Harry Barberian Research Scholarship.

Acknowledgements

This work would not have been possible without the support of my supervisor Dr.

Geoffrey Liu, my advisor Dr. Priya Duggal, as well as my family and friends.

1. Introduction

1.1 Epidemiology

Head and neck cancer is a broad term encompassing a variety of different malignancies occurring within regions of the head and neck including the mouth, nose, para-nasal sinuses, pharynx, skin, or cervical lymph nodes. Amongst primary head and neck cancers, squamous cell carcinoma (SCC) represents the predominant histologic tumor type accounting for over 90% of the cases (Suh, Amelio, Guerrero Urbano, & Tavassoli, 2014). With approximately 650,000 newly diagnosed cases and 350,000 deaths each year, head and neck squamous cell carcinoma (HNSCC) carries a significant burden worldwide (Ferlay et al., 2010). In the United States, there are over 50,000 new cases and 11,000 deaths annually (Jemal, Bray, & Ferlay, 2011).

HNSCC arises from the epithelial lining of the upper aerodigestive tract and can be subcategorized by the involved anatomical sub-site(s). These include the oral cavity, nasopharynx, oropharynx, larynx, and the hypopharynx. Although there exists geographical variations in incidence and mortality for each anatomical sub-site, the oral cavity and oropharynx are the most common sites of disease overall worldwide (Argiris, Karamouzis, Raben, & Ferris, 2008).

HNSCC typically develops in the 6th decade of life. There is largely a male predominance, but the male to female ratio varies from 2:1 to 15:1 depending on anatomical sub-site (Mehanna, Paleri, West, & Nutting, 2010). In general, there has been a trend towards stabilization or decreasing incidence for HNSCC as a whole due to declines in tobacco use (Simard, Torre, & Jemal, 2014). In contrast, an increase in the

global incidence of oropharyngeal cancer has been observed in recent years, particularly in younger age groups (40-50 years old) (Chaturvedi et al., 2013).

Tobacco use and alcohol consumption are the predominant risk factors in approximately 75% of HNSCC cases (Hashibe et al., 2009). The risk associated with tobacco smoking directly correlates with intensity and duration of use. Smoking cessation reduces HNSCC risk, but does not lower it to levels of the never-smoker (Schlect, Franco, Pintos, & Kowalski, 1999). Heavy alcohol consumption is an independent risk factor that exerts a synergistic effect when combined with smoking (Hashibe et al., 2009).

Along with the increase in oropharyngeal carcinoma incidence observed in recent years, the human papillomavirus (HPV) has also emerged as a major etiologic factor in this anatomical sub-site of HNSCCs. HPV infection has been associated with up to 80% of oropharyngeal cancers in the United States, with HPV subtypes 16 and 18 being responsible for the majority of cases (Marur, D'Souza, Westra, & Forastiere, 2010). In other anatomical sub-sites, HPV positivity is much lower. Depending on the method of detection, HPV positivity is associated with less than 10% of oral cavity, laryngeal, and hypopharyngeal tumors (Combes & Franceschi, 2014). Individuals with HPV-positive oropharyngeal carcinoma more often lack the classical risk factors of tobacco and alcohol consumption and have younger ages of disease onset compared to classical HNSCC (Marur, D'Souza, Westra, & Forastiere, 2010).

1.2 Diagnosis and Treatment

Patients with HNSCC present with various signs and symptoms related to the location of the primary tumor and involved structures. These may include pain, mucosal ulceration, odynophagia, dysphagia, dysphonia, hoarseness, otalgia, and neck mass. In

more advanced stages, HNSCC may present with cranial nerve palsies, airway obstruction, and constitutional symptoms including weight loss and fatigue. Biopsy is required for diagnosis. Accurate staging is paramount in guiding therapeutic decision making and involves complete physical examination, endoscopy, and imaging (X-ray, CT, MRI). The seventh edition of the American Joint Committee on Cancer (AJCC) TNM staging system is widely used currently and categorizes disease based on size of the primary tumor, invasion of adjacent structures, cervical lymph node involvement, and distant metastatic spread (Edge & Compton, 2010). Treatment regimen is anatomical sub-site specific and usually involves multiple modalities except in the cases of very early disease. In North America, primary treatment typically consists of either surgical resection or radiotherapy. Adjuvant treatment using a second and/or third modality (i.e. chemotherapy, targeted therapy) is added depending on the extent of disease, patient's age, medical comorbidities, and/or performance status (Argiris et al., 2008).

1.3 Prognosis

Based on Surveillance Epidemiology and End Results (SEER) data, the 5-year survival rate for all sub-sites and stages combined is approximately 65.9% (Pulte & Brenner, 2010). Pathologically confirmed cervical metastases upstages disease to stage III and significantly reduce survival rates by up to 50% (Layland, Sessions, & Lenox, 2005). Distant metastases are late events that represent incurable disease. When detected early and managed in a timely and effective manner, early-staged HNSCC in some anatomical sub-sites may carry a favorable prognosis. The 5-year survival rates based on SEER data is greater than 80% for localized oral cavity and oropharynx carcinoma and approximately 75% for localized laryngeal and nasopharyngeal carcinomas (National

Cancer Institute, 2013; Ou, Zell, Ziogas, & Anton-Culver, 2007). Localized hypopharyngeal carcinoma is associated with the lowest survival rates of less than 60% (Siegel, Ma, Zou, & Jemal, 2014).

HPV positive oropharyngeal carcinoma represents a distinct subgroup of HNSCC in terms of prognosis. Compared to traditional HNSCC associated with smoking and alcohol consumption, HPV-positive oropharyngeal carcinomas are associated with better survival rates and lower risks of recurrence. In this subgroup of patients, significant smoking history negatively impacts prognosis (Ang et al., 2010).

In general, prognostication based on patient and clinical characteristics have proven insufficient alone. Given that individuals with similar disease characteristics and treatments strategies often have variable outcomes, it has become increasingly evident that HNSCC is a complex disease characterized by clinical and biological heterogeneity. Individual genetic variations may underlie this observed heterogeneity (Pai & Westra, 2009). Moreover, for the substantial number of patients presenting with advanced disease, intensive multimodality treatment can cause significant functional impairments (Argiris et al., 2008). As such, there exists a need to identify biomarkers that can strengthen our understanding of the genetic underpinnings of HNSCC and improve our ability to predict outcomes and select patients for individualized therapeutic strategies.

2. Literature Review

2.1 Genetics of HNSCC

The majority of HNSCC occur as sporadic cases. However, inheritable genetic risk factors appear to also play a role. Familial clustering were initially reported in oral and laryngeal cancers, suggesting a genetic component to the disease (Bhaskar, Smith, &

Baughman, 1988; Gencik, Wey, & Muller, 1986; Hara, Ozeki, Shiratsuchi, Tashiro, & Jingu, 1988; Marlowe, 1970). In family- and population-based case-control studies, history of HNSCC in a first-degree relative was associated with a 2-4 fold elevated risk of HNSCC (Copper et al., 1995; Foulkes et al., 1996; Foulkes, Brunet, Kowalski, Narod, & Franco, 1995; Radoï et al., 2013). The impact of genetic predisposition on HNSCC development is further demonstrated in rare cancer susceptibility syndromes including Fanconi's anemia, xeroderma pigmentosum, Li-Fraumeni syndrome, Lynch syndrome, Blooms syndrome, and ataxia-telangiectasia, where incidence of HNSCC is markedly elevated even in the absence of environmental risk factors such as smoking and alcohol consumption (Trizna & Schantz, 1992).

In attempts to identify specific genomic alterations associated with HNSCC, there has been a myriad of studies investigating somatic mutations in HNSCC tumors. From next-generation sequencing studies we have learned that the mutational landscape in HNSCC is diverse with aberrations in genes from a variety of cellular pathways (Agrawal et al., 2011; Lui et al., 2013; Stransky et al., 2011). While mutations in *TP53*, *CDKN2A*, *NOTCH1*, *PIK3CA*, and *HRAS* are the most common, data from The Cancer Genome Atlas (TCGA) has also revealed other novel alterations in the receptor tyrosine kinase family, cell death pathway, NF-kB-mediated survival pathway, and immunity pathway genes (The Cancer Genome Atlas Network, 2015).

2.2 Germline Variations and HNSCC

In comparison to somatic variations that are acquired as tumors develop, germline variations are heritable and non-tissue specific. In the setting of translational research, germline variations can provide an advantage over somatic variations given that they can

be easily detected from sources such as blood or saliva and do not change depending on intra-tumor heterogeneity.

In the past decade, completion of the Human Genome Project and the International HapMap Project coupled with advances in high-throughput technologies have given rise to genome-wide association studies (GWAS) (The International Human Genome Sequencing Consortium, 2002; The International HapMap Consortium, 2003). These studies were traditionally rooted in the Common Disease Common Variant Hypothesis (CVCV), which suggests that susceptibility to common diseases results from common genetic variations (minor allele frequency greater than 1-5%) found in the population of interest (Risch & Merikangas, 1996). These common genetic variations are captured in GWAS in the form of single nucleotide polymorphisms (SNPs), which are inheritable DNA sequence variations of a single base pair. GWAS relies on the principle of linkage disequilibrium (LD), where alleles at two or more genetic loci are non-randomly associated within a population. Because of LD, it is possible to study the associations of genetic variations with phenotypes of interest on a genome-wide scale by genotyping a subset of representative SNPs known as the “tagSNPs”, which are in LD with many other SNPs. Most commonly in a GWAS, unrelated cases are compared to similar individuals without the disease (controls) to determine whether SNP genotype is associated with disease status. With its ability to interrogate across the entire genome in an agnostic fashion, GWAS studies have enriched our understanding of how germline variations contribute to the genetic susceptibility of complex diseases, such as cancer (Stadler et al., 2010).

In terms of HNSCC risk, some evidence from GWAS have emerged, pointing towards significant germline variants within alcohol dehydrogenase genes (rs1573496-*ADH7*, rs1229984-*ADH1B*, rs698-*ADH1C*) as well as near aldehyde dehydrogenase 2 gene (rs4767364) and DNA repair genes *HELQ* and *FAM175A* (rs1494961) (McKay et al., 2011). However, published GWAS studies have been limited by the inclusion of non-HNSCC cases (esophageal cancer) or have focused on only one anatomical sub-site with a very limited sample size (Bhatnagar, Dabholkar, & Saranath, 2012; McKay et al., 2011). As well, the majority of known variants have been examined in the context of HNSCC risk only.

2.3 Germline Variations and HNSCC Survival

A genome-wide approach has not been utilized to examine germline variations in the context of HNSCC prognosis. Although evidence from candidate gene studies has linked polymorphisms in DNA repair, xenobiotic metabolism, cell cycle, and growth factor pathways to survival outcomes in HNSCC, overall there is a paucity of evidence linking genetic variants to HNSCC outcomes, thereby limiting their clinical utility (Hopkins et al., 2008) To date, some associations with HNSCC survival outcomes have been observed with polymorphisms in *XRCC1*, *FGFR4*, and *CCND1*, genes that function in DNA repair, cell growth, and cell cycle control, respectively. While *FGFR4* and *CCND1* polymorphisms have been linked to poor prognosis across several studies, *XRCC1* and *ERCC2* polymorphisms have been associated with improved survival (Azad et al., 2012; da Costa Andrade et al., 2007; Gal, Huang, Chen, Hayes, & Schwartz, 2005; Matthias et al., 1998; Monteiro, Varzim, Pires, Teixeira, & Lopes, 2004; Quintela-Fandino et al., 2006; Streit et al., 2004) (Figure 1). Despite these reported associations,

the studies have generally been limited by several important factors. These include: 1) heterogeneity in study populations, 2) small sample sizes, 3) incomplete documentation of cohort characteristics and outcomes, 4) lack of adequate adjustment for confounders, 5) inadequate correction for multiple-comparisons, and/or 5) non-replication of results. (Hopkins et al., 2008).

Figure 1. Summary of germline variations associated with HNSCC

Function/ pathway	Gene	Variant(s)	Estimate	Study Type
Cell-cycle control	Cyclin D1 (<i>CCND1</i>)	A870G	AHR, 1.38-3.72 (codominant, DFS) ¹ LR, p=0.0095 (OS) ²	Candidate gene
Growth factor pathway	Fibroblast Growth Factor Receptor 4 (<i>FGFR4</i>)	Gly388Arg	AHR, 2.18 (additive, OS) ³ LR, p=0.032 (additive, OS) ⁴	Candidate gene
DNA repair	X-ray repair cross-complementing protein 1 (<i>XRCC1</i>)	Arg399Gln	AHR, 0.68 (additive, OS) ⁵ LR, p=0.0044(OS) ⁶	Candidate gene, gene-set analysis
	Excision repair cross-complementing rodent repair deficiency, complementation group 2 (<i>ERCC2</i>)	Lys751Gln	AHR, 0.80 (additive, DFS) ⁷ LR, p=0.0012 (OS) ⁶	Candidate gene, gene-set analysis

*Abbreviations: OS, overall survival; DFS, disease-free survival; AHR, adjusted hazard ratio; LR, log-rank test

¹Matthias et al., 1998

²Monteiro, Varzim, Pires, Teixeira, & Lopes, 2004

³da Costa Andrade et al., 2007

⁴Streit et al., 2004⁵.Gal, Huang, Chen, Hayes, & Schwartz, 2005

⁶Quintela-Fandino et al., 2006

⁷Azad et al., 2012

3. Study Objective

In order to gain a better understanding of whether germline variations influence HNSCC survival outcomes, we performed a large-scale GWAS in a well-defined

prospectively recruited cohort of HNSCC patients. The main objective of this study was to identify germline variants associated with survival in patients with HNSCC.

Specifically, the aim was to identify common polymorphisms associated with overall survival in patients with primary squamous cell carcinoma of the oral cavity, oropharynx, and hypopharynx.

4. Methods

4.1. Study Enrollment and Databases

Patients with HNSCC diagnosed and treated at the Princess Margaret Cancer Centre in Toronto, Ontario, Canada at the Wharton Head & Neck Centre have been prospectively accrued for an institutional translational study on head and neck cancers since January 2007. The Princess Margaret Cancer Centre is the largest comprehensive cancer centre in Canada, where cancer care is centralized to regional cancer centres with capacity for multidisciplinary care. The Wharton Head & Neck Centre provides care for approximately 800 new patients with head and neck cancer each year.

After histopathologic confirmation of a diagnosis of head and neck cancer, patients were approached to participate in the study during their initial clinic visits. Informed consent for study participation was obtained prior to study enrollment by a study coordinator. All recruited study subjects were 18 years of age or older. Whole blood samples were obtained prior to treatment initiation. A self-administered questionnaire was provided at the time of enrollment to collect demographic and risk factor information including gender, ethnicity, and smoking habits. Additional clinical information including TNM stage, pathology, HPV status, treatment data, outcome data, and duration of follow-up were prospectively collected and stored in the Head and Neck

Anthology of Outcomes Systems-a institutionally developed point-of-care physician-collected electronic data storage system (Wong et al., 2010). HPV status was determined via staining for p16 protein expression in primary tumor tissues using immunohistochemistry. Positive p16 staining is indicative of HPV positive disease. Patients were staged using the American Joint Committee on Cancer TNM classification 7th edition by one or more members of an interdisciplinary team of head and neck surgeons, radiation oncologists, and medical oncologists. Patients treated with primary surgery were staged according to pathological TNM stage of the resected specimen. Those treated with primary radiotherapy were clinically staged based on pre-treatment examination and imaging. Collected survival outcomes were verified using the Ontario Cancer Registry and patients' electronic medical records.

4.2. Ethics Approval

Institutional Ethics approval was obtained from the University Health Network Research Ethics Board for the collection of patient information (clinical and questionnaire) and blood samples as well as subsequent analyses (REB: 07-0521-CE).

4.3. Study Population

For the purpose of this study, patients with histopathologically confirmed primary invasive squamous cell carcinoma of the oral cavity, oropharynx, and hypopharynx were included. Exclusion criteria included refusal of blood collection, non-squamous cell carcinoma pathology, or synchronous cancers. For survival analyses, patients with distant metastases at presentation, palliative treatment intent, treatment at another

institution, and/or had less than two years of follow-up after the date of diagnosis were excluded.

4.4 Data Collection

4.4.1. Demographic and Clinical Information

Demographic information including age, gender, self-reported ethnicity, and smoking status were retrieved from the questionnaire database. TNM stage, HPV status, treatment modality, and outcome data were retrieved from the Head and Neck Anthology of Outcomes Systems.

4.4.2. Blood Collection and DNA Extraction

Blood samples were collected from patients in clinic using Ethylenediaminetetraacetic acid (EDTA) vacutainers. The samples were processed within 2 hours of blood collection (centrifugation at 4 degrees Celsius for 30 minutes at 2000g) at the Princess Margaret Cancer Centre. The plasma and serum of each sample were transferred to cryovials and stored separately. The remaining cellular component at the bottom of the EDTA tubes were aliquoted to cryovials and stored as whole blood at -80 degrees Celsius for up to two years. Prior to genotyping experiments, whole blood samples were thawed at room temperature for 1-2 hours. DNA was then extracted from whole blood samples using 5Primer ArchivePure DNA Blood Kit (Cat# 2300740) according to manufacturers' instructions. DNA concentration and purity were measured using the Nanodrop Spectrophotometer (Thermo Scientific, Wilmington, DE, USA). Samples that met minimal requirements of 50-100 ng/ μ l and total DNA amount of 1-2 μ g were plated on to 96-well plates and sent for genotyping.

4.5 Genotyping

Genotyping was performed using a custom-designed Illumina OncoArray BeadChip at the Center for Inherited Disease Research (CIDR) along with 13 other HNSCC studies that were a part of the International Head and Neck Cancer Epidemiology Consortium (INHANCE)/Genetic Associations and Mechanisms in Oncology (GAME-ON) Consortium. Briefly, the custom OncoArray included a genome-wide backbone of approximately 270,000 tagSNPs as well as over 250,000 additional custom SNPs. Ancestry Informative Markers (AIMs) were included on the array. Custom SNPs selected included those involved in candidate pathways including alcohol metabolism, DNA repair, nicotine addiction, as well as genetic variants known to be of importance in lung squamous cell carcinoma, which share similar risk factors as HNSCC. Candidate variants identified through previous HNSCC GWAS, imputation analyses, and TCGA studies of HNSCC and lung squamous cell carcinoma were also included. HapMap DNA were plated at CIDR in unique positions on each DNA plate and ran with study samples.

4.6 Quality Control

Rigorous quality control (QC) procedures are necessary in GWAS to minimize bias and increase the probability of identifying a true association (Laurie, Doherty, & Mirel, 2010). Prior to the release of genotypes, several technical filters were applied by the genotyping centre at CIDR. These include excluding SNPs with HapMap replicate errors, samples with high duplicate errors, and samples with contaminated DNA. Genotyping calls were further reviewed and cleaned by an independent group within INHANCE with expertise in GWAS genotype calling and quality control. A systematic

pipeline of quality control procedures was then conducted on the OncoArray genotype data. The quality control procedures were implemented in PLINK version 1.09 and R studio Version 0.98.1103. Principal Components Analysis (PCA) was performed using EIGENSOFT version 6.0.1. Both sample-based and marker-based quality control were performed as follows:

4.6.1 Gender Check

Sex inconsistencies were identified by comparing the reported sex of individuals against their predicted sex as determined by their X chromosome heterozygosity rate calculated in PLINK. Individuals with sex chromosome abnormalities were identified. If the identified sex chromosome abnormality had no known influence on HNSCC prognosis, the individuals were not excluded from downstream analyses. Individuals with unresolved sex-discrepancies were excluded.

4.6.2 Relatedness

Cryptic relatedness can increase the type 1 error rate. Pairwise familial relationships were estimated using a method of moments procedure implemented in PLINK. Identify-by-descent (IBD) coefficients (Z_0 , Z_1 , and Z_2), representing the probability of sharing 0, 1, or 2 alleles identical by descent, were estimated. The probability of sharing 0 (Z_0) alleles IBD was then plotted against the probability of sharing 1 allele (Z_1) IBD for each pair. One individual from each pair with proportion IBD greater than 0.120 (more closely related than first cousins) was excluded.

4.6.3 Genotype call rates by individual and SNP

Sample call rate refers to the proportion of SNPs that were successfully genotyped per sample. Samples where a significant proportion of SNPs failed genotyping may contain poor quality DNA that can give rise to erroneous genotype calls. Samples with call rates less than 98% were excluded. Similarly, marker call rate, which is the proportion of samples with a successful genotype call for each marker, was assessed as an indicator of marker quality. SNPs with call rates below 98% were excluded.

4.6.4 Minor Allele Frequency

Distribution of minor allele frequency (MAF) across SNPs were examined. To ensure adequate power and minimize false positives, SNPs with $MAF < 0.05$ were excluded from survival analyses.

4.6.5 Batch Effects

Genotyping for the study was completed across a total of thirteen 96-well plates. As such, each plate was examined and compared in terms of sample call rate and MAF to assess for potential differences by plate.

4.6.6 Population Substructure

The presence of population stratification is a major source of confounding in GWAS. False results may be observed due to differences in ancestry rather than true association between genotype and outcome of interest. Principal components analysis (PCAs) using a panel of 1847 ancestry informative markers was used to assess for population stratification. Study genotype data was first examined alone. Study genotypes were then merged with HapMap phase 3 data in order to further define population

identities. Population outliers (exceeding 6 standard deviations) as identified along the top 10 principal components using EIGENSOFT were excluded from further analyses.

4.6.7 Heterozygosity

Heterozygosity rates that significantly deviate from the average may be indicative of possible sample contamination, inbreeding, and/or poor data quality. Heterozygosity rate was estimated using PLINK, which computed the observed and expected autosomal homozygous genotype counts for each sample. To obtain the heterozygosity rate, the number of observed number of homozygotes was subtracted from the number of non-missing autosomal genotypes and divided by the number of non-missing autosomal genotypes. The heterozygosity rate was then plotted for each sample to identify outliers that were more than 6 standard deviations away from the mean.

4.6.8 Hardy-Weinberg Equilibrium

Checking whether each marker was in Hardy-Weinberg Equilibrium (HWE) was the final step in the quality control procedures. Under Hardy-Weinberg assumptions, allele and genotype frequencies in a population remain constant from one generation to the next. SNPs with significant deviations from HWE may be suggestive of genotyping errors. However, deviations from HWE can also be observed when a true association is present. As such, the HWE p value cutoff was set to 10^{-6} in this study.

4.7 Statistical Analysis

The outcome of interest was overall survival, which was defined as the time interval between the date of diagnosis and the date of death from any cause. Duration of follow-up was calculated from the date of diagnosis until date of last known follow-up or

contact. Patients known to be alive were censored at the time of last contact, up to February 28, 2015. Cox proportional hazards models were used to assess the association between SNPs and overall survival. Autosomal SNPs that met quality control criteria were included in the genome-wide analysis. Multivariate models adjusting for known clinical confounders were utilized. Multicollinearity of variables was assessed using variance inflation factor. The proportional hazards assumption was assessed for included variables. Heteroscedasticity-consistent standard errors were used in the Cox multivariate models. Three genetic models of inheritance (additive, dominant, and recessive) were assessed and the model of best fit was selected for analyses. Top SNPs meeting the threshold for “suggestive” association (p value $<10^{-5}$) were further investigated. Kaplan-Meier survival estimates were plotted and the log rank test was used to compare survival times across different genotypes. Regional coverage around significant SNPs and LD between SNPs were assessed using LocusZoom version 1.1. Exploratory subgroup analyses based on anatomical sub-site were conducted.

5. Results

5.1 Quality Control

Detailed results of quality control procedures are shown in the supplementary information section. An overview of samples excluded during quality control procedures are shown in Figure 2. After completing sample-based quality control procedures, 47 individuals were excluded. Twelve samples were excluded by CIDR due to poor performance. Eleven individuals were flagged during gender check. Of these, 3 were identified as database errors, where the gender by genotype was correct but was erroneously recorded as the opposite sex in the clinical database. These errors were

subsequently corrected and these individuals remained included in the analyses. Four individuals were identified as having sex chromosome abnormalities. Specifically, 3 patients were identified as XXY and one as XX/XO. Given that these sex chromosome abnormalities were not known to impact HNSCC survival, these individuals remained in subsequent analyses. Two individuals who underwent bone marrow transplantation had genotyped genders matching that of their donors'. These individuals were excluded along with two additional individuals with unresolved gender inconsistency. Six individuals were excluded due to cryptic relatedness. Overall, the per sample call rates were high with a mean call rate of 98.7%. Twenty samples with call rates less than or equal to 98% were excluded. After exclusion, the mean per sample call rate was 99.9%. During analysis of population substructure, 4 individuals were excluded given that they were 6 standard deviations away from the top 10 PCA means. One individual was removed due to an extreme heterozygosity rate.

Figure 2. Summary of sample quality control

Number of samples before sample quality control: 1,145	
<i>Reason for Exclusion</i>	<i>Excluded</i>
Poor performance (CIDR)	12
Unresolved sex inconsistency	2
Bone marrow transplant	2
Cryptic relatedness	6
Sample call rate <98%	20
Population outliers	4
Heterozygosity rate >6 S.D. from mean	1
	Total=47
Number of Samples After Sample Quality Control: 1,098	

In total, genotyping for 533,631 SNPs was attempted including 517,820 autosomal SNPs, 15,258 X chromosome SNPs, 397 Y chromosome SNPs, 6 XY chromosome SNPs, and 150 mitochondrial SNPs. The mean per SNP call rate was

98.7%. A total of 19,071 SNPs not meeting the 98% call rate cut off and 748 duplicate SNPs were excluded. A total of 35,827 monomorphic SNPs were excluded. The distribution of minor allele frequencies across SNPs are shown in Figure 5 (Supplementary Information). After exclusions, the mean per SNP call rate was 99.9% and the mean minor allele frequency was 0.2013. No significant batch effects were observed across the 13 genotyping plates. Overall, 55,646 variants were excluded after marker-based quality control procedures, with 477,985 variants remaining for further analyses. A summary of SNP-based quality control procedures is shown in Figure 3.

Figure 3. Summary of SNP quality control

Number of SNPs before QC: 533,631	
<i>Reason for Exclusion</i>	<i>Excluded</i>
SNP call rate <98%	19,071
Duplicate SNPs	748
Monomorphic SNPs	35,827
	Total=55,646
Number of Samples After Sample QC: 477,985	

5.2 Patient Characteristics

Between 2007-2014, 1178 patients with newly diagnosed squamous cell carcinoma of the oral cavity, oropharynx, and hypopharynx were recruited into the study. Thirty-three individuals were excluded due to ineligible anatomical sub-site or blood samples. After completing quality control procedures, 1098 patients remained. Prior to survival analysis, 172 individuals were excluded due to inadequate follow-up time (less than 2 years) and 42 were excluded due to treatment intent and/or treatment elsewhere. A total of 884 individuals were eligible for survival analyses. Their demographic and clinicopathologic characteristics are shown in Table 1. See Supplementary Information Figure 1 for an overview of included/excluded samples.

Table 1. Characteristics of patients eligible for survival analysis (N=884)

	N	%
Age at diagnosis		
Mean	61.2+/- 11.0 years	
Range	27.0-91.0 years	
Gender		
Male	669	75.7
Female	215	24.3
Self-Reported Ethnicity		
Caucasian	727	86.0
South Asian	50	5.9
East Asian	33	3.9
Black	9	1.0
Other	26	2.9
NA	39	4.4
Anatomical Sub-site		
Oral Cavity	385	43.6
Oropharynx	451	51.0
Hypopharynx	48	5.4
HPV status		
Positive	316	35.7
Negative	114	12.9
Unknown	454	51.4
Primary Treatment		
Surgery	372	42.1
Radiotherapy+/- chemotherapy	512	57.9
Vital Status		
Alive	686	77.6
Dead	198	22.4
Duration of follow-up		
Mean	3.1 +/-1.8 years	
Range	0.1-7.6 years	

*NA=No self-reported ethnicity information

The majority of patients were male and self-identified Caucasian. Using principal components, 768 individuals of European descent were further defined based on their similarity to CEU and TSI HapMap populations. Other populations, namely Indian, East Asian, African, and Others were also identified using reference HapMap populations. The most common anatomical sub-site was oropharynx followed by oral cavity. Hypopharyngeal cancer was the diagnosis in a small proportion of patients only. Within

the oropharyngeal sub-site, the majority of patients had p16 positive disease attributable to HPV infection. Radiotherapy was more common than surgical resection as the primary treatment modality. The average duration of follow-up was approximately 3 years. The total number of deaths in the study was 198. The 1, 3, and 5-year survival rates were 91.3%, 78.2%, and 73.3%, respectively. On univariate analyses, age, overall TNM stage, anatomical sub-site, and treatment modality were significantly associated with overall survival. Gender was not significantly associated with survival (results not shown).

5.3 Germline Variants Associated with HNSCC Survival

To ensure adequate power, only SNPs with MAF equal to or greater than 0.05 were considered for survival analyses. None of the top SNPs had allele frequencies that significantly deviated from HWE. A Cox proportional hazards model adjusting for age, overall TNM stage, and anatomical sub-site under the additive model of genetic inheritance was fitted to assess the association between SNPs and overall survival. Treatment was highly collinear with anatomical sub-site and therefore was not included in the multivariate model. The primary analysis was performed within patients of European descent (as identified by PCA) only, given that this was the predominant ethnic group (87% of patients eligible for survival analyses). The Quantile-Quantile (Q-Q) plot did not demonstrate systemic inflation of the test statistic ($\lambda_{\text{median}}=0.994$) (Figure 4). The Manhattan plot for overall survival in European patients is shown in Figure 5. Although no single SNP reached the level of genome-wide significance at $p < 5 \times 10^{-8}$, 3 identified variants were suggestive of significance ($p < 1 \times 10^{-6}$).

Figure 4. Q-Q plot of GWAS for overall survival in European HNSCC patients.

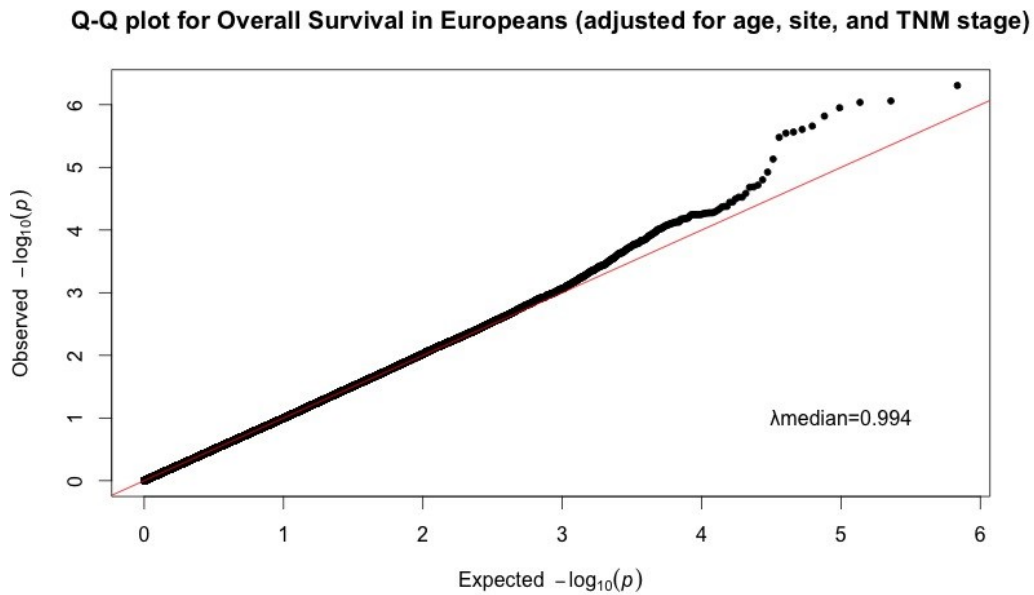
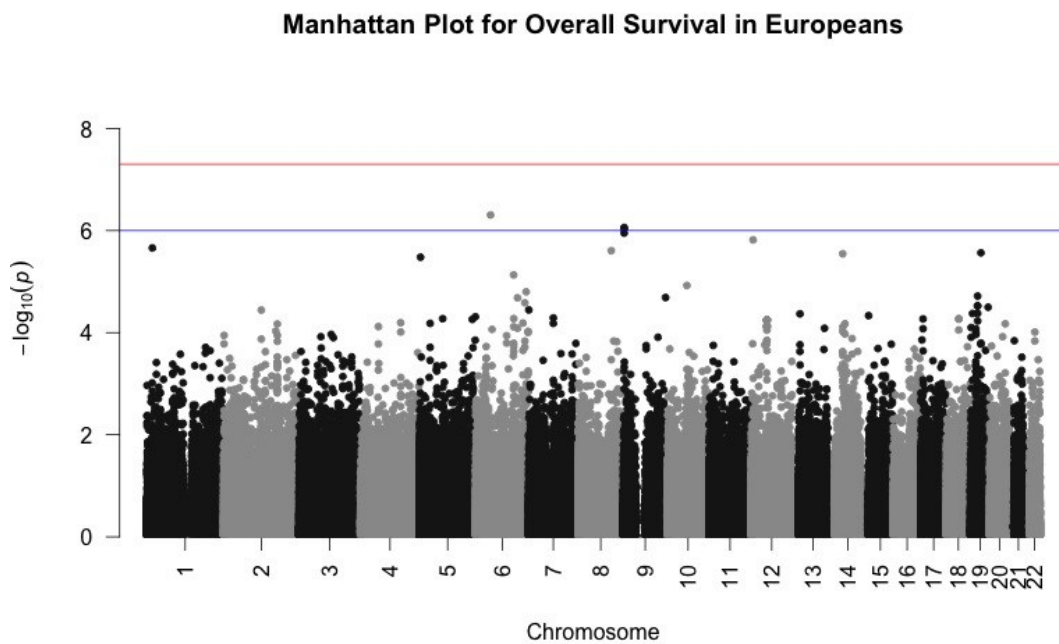


Figure 5. Manhattan plot of GWAS for overall survival in European HNSCC patients



*In Figures 4 and 5, associations expressed as $-\log_{10}(p)$. P values were two-sided and were from multivariable Cox proportional hazards models adjusted for age, site, and overall TNM stage.

The SNPs most significantly associated with overall survival in Europeans are summarized in Table 2. All top SNPs were found within annotated genes. The Kaplan-Meier plots for top SNPs are shown in Figure 6.

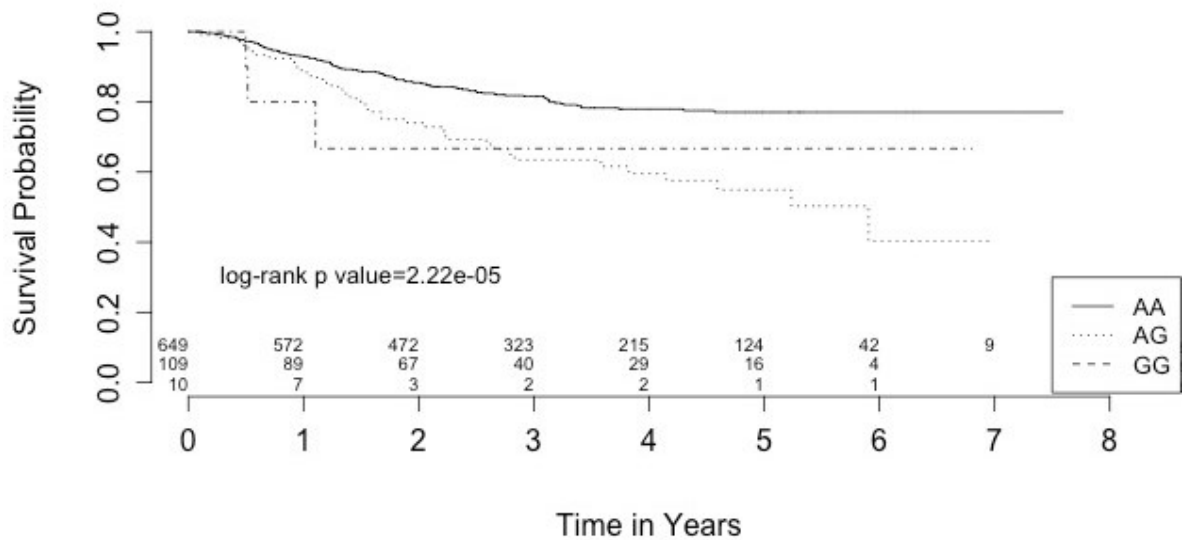
Table 2. Significant SNPs associated with overall survival in European HNSCC patients (N=768)

Ch	Band	SNP	Position	Gene	Risk Allele	MAF	HR (95% CI)	P-value
6	6p12.3	rs1974051	47916225	<i>PTCHD4</i>	G	0.08	2.09(1.57-2.79)	4.94x10 ⁻⁷
9	9p24.2	rs10812227	2548556	<i>VLDLR-AS1</i>	T	0.14	0.32(0.20-0.50)	8.71x10 ⁻⁷
9	9p24.2	rs12237653	2551654	<i>VLDLR-AS1</i>	C	0.14	0.32(0.20-0.50)	9.18x10 ⁻⁷

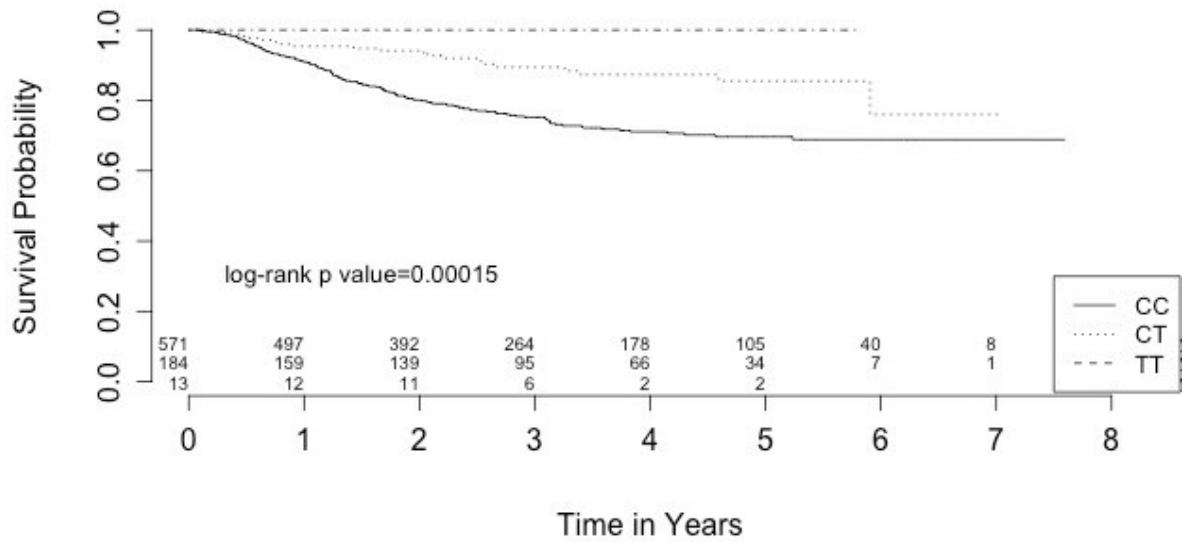
*Abbreviations: Ch, Chromosome; MAF, minor allele frequency; HR, hazard ratio; CI, confidence interval

*SNP names based on dbSNP137

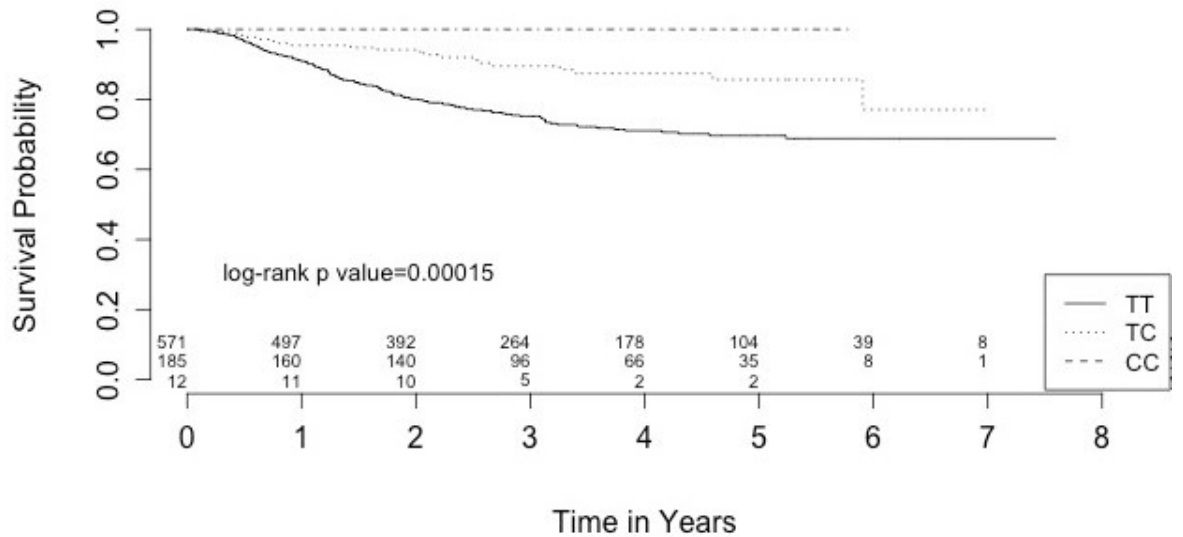
Figure 6. Unadjusted Kaplan-Meier curves for overall survival by genotype
a. rs1974051



b. rs10812227



c. rs12237653



The top-ranking SNP (rs1974051) was located within the patched domain containing 4 (*PTCHD4*) gene at the 6p12.3 locus. Each minor allele was associated with a hazard ratio (HR) of 2.09 (95% CI, 1.57-2.79). The other 2 top-ranking SNPs (rs10812227 and rs12237653), along with one additional SNP (rs4741732) that was trending towards significance, were found within the 9p24.2 region. These SNPs showed

a protective effect in terms of overall survival (HR, 0.32; 95% CI, 0.20-0.50). They were in high LD with one another and were clustered within the very low density lipoprotein receptor antisense RNA 1 gene (*VLDLR-AS1*) (Supplementary Information Figure 10).

There were only 116 non-European patients and the number of patients within each non-European subgroup was small (63 South Asian, 27 East Asians, 9 African, and 17 others/admixed). As such, a trans-ethnic meta-analysis was not performed.

5.4 Sensitivity Analyses Based on Anatomical Sub-site

Etiology and survival outcomes in HNSCC significantly vary by anatomical sub-site. As such, exploratory subgroup analyses were conducted within the oropharynx and oral cavity sub-sites. The sample size of the hypopharynx subgroup was too small to yield reliable estimates.

Within the oropharynx, in addition to age and TNM stage, Cox proportional hazard models were also adjusted for HPV status as determined by p16 positivity and smoking status. An intergenic variant (rs7862541) located within the 9p24.1 region met the genome-wide significance threshold ($p=2.44 \times 10^{-8}$). Six additional variants were suggestive of significance as shown in Table 3.

Table 3. Significant SNPs in the oropharynx sub-site in European HNSCC patients (N=376)

Ch	Band	SNP	Position	Gene	Risk Allele	MAF	HR (95% CI)	P-value
9	9p24.1	rs7862541	5604769	intergenic	T	0.10	2.98(2.03-4.38)	2.44×10^{-8}
4	4q32.2	rs75965042	164037795	<i>NAF1</i>	C	0.06	3.82(2.34-6.23)	7.43×10^{-8}
4	4q32.2	rs17574945	164056625	<i>NAF1</i>	T	0.06	3.82(2.34-6.23)	7.43×10^{-8}
1	1p34.1	rs6667593	45131401	<i>TMEM53</i>	C	0.06	2.65(1.83-3.82)	2.29×10^{-7}
1	1q21.3	rs7523883	153214867	intergenic	C	0.34	2.24(1.64-3.05)	3.93×10^{-7}
4	4q32.3	rs62335075	164155332	intergenic	T	0.20	2.41(1.70-3.41)	6.85×10^{-7}
17	17q21.2	rs72829884	46681084	<i>HOXB6</i>	C	0.07	3.36(2.08-5.43)	7.62×10^{-7}

*Abbreviations: Ch, Chromosome; MAF, minor allele frequency; HR, hazard ratio; CI, confidence interval

*SNP names based on dbSNP137

*Cox-proportional hazard model adjusted for age, stage, p16 status, and smoking status

Within the oral cavity subgroup, the top variant reaching genome-wide significance was an intergenic variant located upstream of the patched 1 (*PTCH1*) gene ($p=2.27 \times 10^{-10}$). Three additional SNPs achieved suggestive significance (Table 4).

Table 4. Significant SNPs in the oral cavity sub-site in European HNSCC patients (N=304)

Ch	Band	SNP	Position	Gene	Risk Allele	MAF	HR (95% CI)	P-value
9	9q22.3	rs55957307	98337392	intergenic	C	0.07	3.86(2.54-5.85)	2.27×10^{-10}
2	2q32.3	rs79913292	192698793	intergenic	G	0.05	3.31(2.08-5.29)	5.05×10^{-7}
10	10q21.2	rs73255291	61849599	<i>ANK3</i>	A	0.06	2.86(1.89-4.32)	5.94×10^{-7}
6	6q12.3	rs1974051	47916225	<i>PTCHD4</i>	G	0.08	2.74(1.83-4.09)	8.55×10^{-7}

*Abbreviations: Ch, Chromosome; MAF, minor allele frequency; HR, hazard ratio; CI, confidence interval

*SNP names based on dbSNP137

*Cox-proportional hazard model adjusted for age, stage

6. Discussion

6.1 Germline Variations and Cancer Survival

While it has been established that germline variations are associated with complex disease risk, a growing body of evidence suggests that they may also be important in improving our understanding of as well as our ability to predict disease outcomes. In cancer, germline variations may impact disease progression and prognosis by affecting hosts' response to treatment as well as by influencing tumors' somatic and epigenetic landscapes (Coate et al., 2010; Dworkin et al., 2010; Pujana, 2014). Utilizing candidate gene and genome-wide approaches, germline variants have been associated with treatment response and/or resistance, relapse and metastasis susceptibility, as well as survival time in a variety of solid tumors and hematologic malignancies. Specifically, GWAS has been utilized to assess germline variations' association with survival outcomes in lung, breast, ovarian, pancreatic, colorectal cancers, and acute lymphoblastic leukemia (Braun et al., 2013; Han et al., 2014; Hu et al., 2012; Innocenti et al., 2012;

Rafiq et al., 2013; Shu et al., 2012; Tang et al., 2015; C. Wu et al., 2014; X. Wu et al., 2013; L. Xu et al., 2012; W. Xu et al., 2015; Yang et al., 2012).

To our knowledge, this is the first study to utilize a genome-wide approach to investigate germline variants' association with survival outcomes in HNSCC. Out of over 530,000 SNPs examined, although no SNPs reached the genome-wide significance threshold of 5×10^{-8} , we identified three common germline variants that may influence survival outcomes in HNSCC patients.

6.2 Germline Variants Associated with HNSCC Survival

The most statistically significant SNP (rs1974051) from the Europeans only all sub-sites analysis was located within an intronic region of the *PTCHD4* gene. The allele G, present in approximately 8% of European patients, was associated with a two-fold decrease in overall survival. *PTCHD4*, also known as *PTCH53*, is a homolog of *PTCH1* - the receptor for Sonic Hedgehog (SHH). Although the exact function of *PTCHD4* has not been well characterized, recent evidence suggests that it may function similarly to other patched family proteins and act as a suppressor of Hedgehog signaling (Chung, Larsen, Chen, & Bunz, 2014).

The Hedgehog pathway regulates key human development processes including cellular differentiation and proliferation. Inappropriate activation of this pathway has been linked to tumorigenesis in a variety of human cancers (Harris, Samant, & Shevde, 2011; Rubin & de Sauvage, 2006). In HNSCC, overexpression of hedgehog pathway components have been observed (Dimitrova et al., 2013). It has been suggested that aberrant hedgehog signaling may play an important role in the pathogenesis of HPV negative tumors (Fertig et al., 2013; Gan et al., 2014). Our sensitivity analysis by

anatomical sub-site's finding of a significant variant near the *PTCH1* gene within the oral cavity, a predominantly HPV-negative sub-site, was in keeping with this hypothesis. In addition, Hedgehog activation has been associated with resistance to radiotherapy, epithelial to mesenchymal transition, and rapid tumor repopulation following treatment (Gan et al., 2014; Ma et al., 2013; Ohta et al., 2009; Sims-Mourtada et al., 2006). It was previously found in the landmark RTOG 90-03 trial that individuals with high expression of *GLI1*, a downstream target of Hedgehog activation, had decreased overall survival, poor local control, and high rates of distant metastases (Chung et al., 2011). Inhibition of the Hedgehog pathway by targeted agents has also shown promise in enhancing the antitumor effects of chemoradiation (Gan et al., 2014; Mozet, Stoeck, Dimitrova, Dietz, & Wichmann, 2013). Further studies are required to elucidate the function of *PTCHD4* and *PTCH1* in the context of HNSCC and to uncover the mechanisms through which germline variations in the Hedgehog pathway may result in differences in HNSCC progression and outcomes. This study contributes to the growing body of literature from in vitro, in vivo, and epidemiologic studies demonstrating the importance of the Hedgehog pathway as a potential oncogenic driver and therapeutic target in HNSCC.

In the overall analysis, the other two variants (rs10812227 and rs12237653) suggestive of a significant association were also intronic variants. They were in high LD with one another and were both situated within *VLDLR-AS1*. In European patients, these variants were relatively common with a MAF of 0.14 for both. The variants were located within a chromosomal region that has long been linked to HNSCC. Loss of heterozygosity within the 9p21-24 region is one of the most frequently observed chromosomal alterations in HNSCC, which has been associated with poor prognosis

(Coon et al., 2004; Gollin, 2001; Rybicki, 2003). In addition to the well-known *CDKN2A* gene located at 9p21 encoding for the p16 protein, the region has also been suspected to harbor additional HNSCC tumor suppressor genes (Beder et al., 2003; Gunduz et al., 2009). The *SMARCA2* gene, located approximately 350kb downstream from rs10812227 and rs12237653, is one of such putative tumor suppressors that has been examined by our group and others (Gunduz et al., 2009; Wang et al., 2013).

VLDLR-AS1, located at 9p24.2, is a non-protein coding gene that encodes for a long non-coding RNA (lncRNA). The RNA is transcribed in the antisense direction relative to the protein-coding *VLDLR* gene, however its exact function remains unknown. While non-protein coding DNA was once considered genomic “junk” with no functional significance, it has become increasingly apparent in recent years that they serve important regulatory functions (Amaral, Dinger, Mercer, & Mattick, 2008; Cheetham, Gruhl, Mattick, & Dinger, 2013; Dunham et al., 2012; Kapranov, Willingham, & Gingeras, 2007). In particular, lncRNAs, which are noncoding RNAs of greater than 200bp in length, have been shown to play integral roles in cancer development and progression. LncRNAs can be found in intergenic regions, within introns of protein-coding genes, or in sense or antisense orientation to protein-coding genes (Cheetham et al., 2013; Mercer, Dinger, & Mattick, 2009). Antisense non-coding RNAs, such as *VLDLR-AS1*, are a type of lncRNA that can regulate the expression of their neighboring genes in *cis* or affect more distant loci *in trans* (Villegas & Zaphiropoulos, 2015). Although some studies have linked alterations in *VLDLR* to cancer including the presence of *VLDLR* somatic mutations in 6% of HNSCCs in TCGA as well as an association between overexpression of a *VLDLR* splice variant and cancer metastasis, the level of evidence remains low

overall (Cerami et al., 2012; He et al., 2010; He, Lu, & Guo, 2013). Given that antisense RNA can modulate chromatin structure and recruit epigenetic effectors to specific genomic loci, it is possible that the candidate germline variants in *VLDLR-ASI* may functionally impact neighboring gene(s) with known significance in HNSCC, such as *SMARCA2* (Mercer & Mattick, 2013). Alternatively, it is also possible that the identified polymorphisms in *VLDLR-ASI* do not have functional significance but served as markers for casual variant(s) in LD at another nearby gene that directly impact HNSCC survival.

6.3 Limitations

This study identified novel germline variants that may impact survival outcomes in HNSCC patients. However, there were several limitations. Despite having a moderately large sample size of HNSCC patients, a reason for the lack of associations reaching genome-wide significance may still be limited power. In general, there is a paucity of prospectively recruited cancer cohorts with large sample sizes and comprehensive demographic and survival information, thereby limiting the progress of GWAS on cancer survival. As such, the formation of collaborative consortia amongst multiple institutions is key to accruing large enough sample sizes to achieve adequate power. The next stage of this study will involve a combined analysis of additional HNSCC cohorts in the INHANCE consortium recruited from various centers within the United States and Europe. The presence of heterogeneity in terms of anatomical sub-sites also contributed to the limited power. Our sensitivity analyses suggested that different anatomical sub-sites of HNSCC may be characterized by distinct germline genetic profiles. Further studies based on single sub-site and/or HPV positivity will likely

improve our ability to detect robust associations with survival outcomes compared to studies that include all HNSCC patients.

Validations of our findings in other independent HNSCC cohorts with European patients is needed. Examination of non-European cohorts and follow-up functional studies are also needed. Similar to our study, the majority of GWAS hits to date lie within intergenic or intronic regions (Hindorff et al, 2015). It has been hypothesized that these types of variants indirectly influence gene regulation instead of directly disrupt protein products (Edwards, Beesley, French, & Dunning, 2013). In order to examine the top candidate variants' functional significance, we plan to perform fine mapping of the identified loci via imputation analyses, in-silico analyses, and expression studies (i.e. luciferase reporter assays, gain/loss of function studies). Future direction also includes pathway-based analyses, examination of cancer-specific survival, and trans-ethnic analyses.

7. Conclusions

Employing a genome-wide approach, we found preliminary evidence suggesting that germline variations may influence prognosis in patients with primary HNSCC. However, further studies are required to validate these findings and to uncover the mechanisms underlying their functional significance. With its ease of detection, germline variants can be used as valuable biomarkers to guide prognostication and treatment selection in the clinical setting. Since multi-modality treatment of HNSCC often results in significant morbidity, treatment de-escalation in patients with favorable disease features is an area of ongoing research. Findings from our study, particularly the discovery of protective variants rs10812227 and rs12237653, may help to identify such a

subgroup of patients. In order to ultimately realize personal cancer medicine in HNSCC, it is likely that the integration of germline, somatic, histopathological, and clinical data will be required.

References

- Agrawal, N., Frederick, M. J., Pickering, C. R., Bettegowda, C., Chang, K., Li, R. J., ... Myers, J. N. (2011). Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science*, 333(6046), 1154–1157. doi:10.1126/science.1206923
- Amaral, P. P., Dinger, M. E., Mercer, T. R., & Mattick, J. S. (2008). The eukaryotic genome as an RNA machine. *Science*, 319, 1787–1789. doi:10.1126/science.1155472
- Ang, K. K., Harris, J., Wheeler, R., Weber, R., Rosenthal, D. I., Nguyen-Tân, P. F., ... Gillison, M. L. (2010). Human papillomavirus and survival of patients with oropharyngeal cancer. *The New England Journal of Medicine*, 363, 24–35. doi:10.1056/NEJMoa0912217
- Argiris, A., Karamouzis, M. V., Raben, D., & Ferris, R. L. (2008). Seminar Head and neck cancer. *The Lancet*, 371, 1695–1709. doi:10.1016/S0140-6736(08)60728-X
- Azad, A. K., Bairati, I., Samson, E., Cheng, D., Mirshams, M., Qiu, X., ... Liu, G. (2012). Validation of genetic sequence variants as prognostic factors in early-stage head and neck squamous cell cancer survival. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, 18(1), 196–206. doi:10.1158/1078-0432.CCR-11-1759
- Beder, L. B., Gunduz, M., Ouchida, M., Fukushima, K., Gunduz, E., Ito, S., ... Shimizu, K. (2003). Genome-wide analyses on loss of heterozygosity in head and neck squamous cell carcinomas. *Laboratory Investigation; a Journal of Technical Methods and Pathology*, 83(1), 99–105. doi:10.1097/01.LAB.0000047489.26246.E1
- Bhaskar, P. B., Smith, R. G., & Baughman, R. A. (1988). Oral squamous cell carcinoma in identical twins: report of a case. *Journal of Oral Maxillofacial Surgery*, 46, 1096–1098.
- Bhatnagar, R., Dabholkar, J., & Saranath, D. (2012). Genome-wide disease association study in chewing tobacco associated oral cancers. *Oral Oncology*, 48(9), 831–835. doi:10.1016/j.oraloncology.2012.03.007
- Braun, R., Finney, R., Yan, C., Chen, Q. R., Hu, Y., Edmonson, M., ... Buetow, K. (2013). Discovery Analysis of TCGA Data Reveals Association between Germline Genotype and Survival in Ovarian Cancer Patients. *PLoS ONE*, 8(3). doi:10.1371/journal.pone.0055037
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., ... Schultz, N. (2012). The cBio Cancer Genomics Portal: An open platform for exploring

- multidimensional cancer genomics data. *Cancer Discovery*, 2(May), 401–404. doi:10.1158/2159-8290.CD-12-0095
- Chaturvedi, A. K., Anderson, W. F., Lortet-Tieulent, J., Curado, M. P., Ferlay, J., Franceschi, S., ... Gillison, M. L. (2013). Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, 31(36), 4550–4559. doi:10.1200/JCO.2013.50.3870
- Cheetham, S. W., Gruhl, F., Mattick, J. S., & Dinger, M. E. (2013). Long noncoding RNAs and the genetics of cancer. *British Journal of Cancer*, 108(April), 2419–2425. doi:10.1038/bjc.2013.233
- Chung, C. H., Dignam, J. J., Hammond, M. E., Klimowicz, A. C., Petrillo, S. K., Magliocco, A., ... Ang, K. K. (2011). Glioma-associated oncogene family zinc finger 1 expression and metastasis in patients with head and neck squamous cell carcinoma treated with radiation therapy (RTOG 9003). *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, 29(10), 1326–1334. doi:10.1200/JCO.2010.32.3295
- Chung, J. H., Larsen, a. R., Chen, E., & Bunz, F. (2014). A PTCH1 Homolog Transcriptionally Activated by p53 Suppresses Hedgehog Signaling. *Journal of Biological Chemistry*, 289(47), 33020–33031. doi:10.1074/jbc.M114.597203
- Coate, L., Cuffe, S., Horgan, A., Hung, R. J., Christiani, D., & Liu, G. (2010). Germline genetic variation, cancer outcome, and pharmacogenetics. *Journal of Clinical Oncology*, 28(26), 4029–4037. doi:10.1200/JCO.2009.27.2336
- Combes, J.-D., & Franceschi, S. (2014). Role of human papillomavirus in non-oropharyngeal head and neck cancers. *Oral Oncology*, 50(5), 370–379. doi:10.1016/j.oraloncology.2013.11.004
- Coon, S. W., Saveria, A. T., Zarbo, R. J., Benninger, M. S., Chase, G. a, Rybicki, B. a, & Van Dyke, D. L. (2004). Prognostic implications of loss of heterozygosity at 8p21 and 9p21 in head and neck squamous cell carcinoma. *International Journal of Cancer. Journal International Du Cancer*, 111(2), 206–212. doi:10.1002/ijc.20254
- Copper, M. P., Jovanovic, A., Nauta, J. J. P., Braakhuis, B. J. M., De Vries, N., Van Der Waal, I., & Snow, G. B. (1995). Role of genetic factors in the etiology of squamous cell carcinoma of the head and neck. *Archives of Otolaryngology, Head & Neck Surgery*, 121(2), 157–160.
- Da Costa Andrade, V. C., Parise, O., Hors, C. P., de Melo Martins, P. C., Silva, A. P., & Garicochea, B. (2007). The fibroblast growth factor receptor 4 (FGFR4) Arg388 allele correlates with survival in head and neck squamous cell carcinoma.

- Experimental and Molecular Pathology*, 82, 53–57.
doi:10.1016/j.yexmp.2006.05.003
- Dimitrova, K., Stoehr, M., Dehghani, F., Dietz, A., Wichmann, G., Bertolini, J., & Mozet, C. (2013). Overexpression of the Hedgehog Signalling Pathway in Head and Neck Squamous Cell Carcinoma. *Onkologie*, 36. doi:10.1159/000350322
- Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. a., Doyle, F., ... Birney, E. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, 57–74. doi:10.1038/nature11247
- Dworkin, A. M., Ridd, K., Bautista, D., Allain, D. C., Iwenofu, O. H., Roy, R., ... Toland, A. E. (2010). Germline variation controls the architecture of somatic alterations in tumors. *PLoS Genetics*, 6(9), 1–9. doi:10.1371/journal.pgen.1001136
- Edge, S. B., & Compton, C. C. (2010). The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Annals of Surgical Oncology*, 17, 1471–1474. doi:10.1245/s10434-010-0985-4
- Edwards, S. L., Beesley, J., French, J. D., & Dunning, M. (2013). Beyond GWASs: Illuminating the dark road from association to function. *American Journal of Human Genetics*, 93, 779–797. doi:10.1016/j.ajhg.2013.10.012
- Ferlay, J., Shin, H.-R., Bray, F., Forman, D., Mathers, C., & Parkin, D. M. (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer. Journal International Du Cancer*, 127(12), 2893–2917. doi:10.1002/ijc.25516
- Fertig, E. J., Markovic, A., Danilova, L. V., Gaykalova, D. a., Cope, L., Chung, C. H., ... Califano, J. a. (2013). Preferential activation of the hedgehog pathway by epigenetic modulations in HPV negative HNSCC identified with meta-pathway analysis. *PLoS ONE*, 8(11). doi:10.1371/journal.pone.0078127
- Foulkes, W. D., Brunet, J. S., Kowalski, L. P., Narod, S. a, & Franco, E. L. (1995). Family history of cancer is a risk factor for squamous cell carcinoma of the head and neck in Brazil: a case-control study. *International Journal of Cancer. Journal International Du Cancer*, 63, 769–773.
- Foulkes, W. D., Brunet, J. S., Sieh, W., Black, M. J., Shenouda, G., & Narod, S. a. (1996). Familial risks of squamous cell carcinoma of the head and neck: retrospective case-control study. *BMJ (Clinical Research Ed.)*, 313, 716–721.
- Gal, T. J., Huang, W.-Y., Chen, C., Hayes, R. B., & Schwartz, S. M. (2005). DNA repair gene polymorphisms and risk of second primary neoplasms and mortality in oral cancer patients. *The Laryngoscope*, 115(December), 2221–2231. doi:10.1097/01.mlg.0000183736.96004.f7

- Gan, G. N., Eagles, J., Keysar, S. B., Wang, G., Glogowska, M. J., Altunbas, C., ... Jimeno, a. (2014). Hedgehog Signaling Drives Radioresistance and Stroma-Driven Tumor Repopulation in Head and Neck Squamous Cancers. *Cancer Research*, 74(23), 7024–7036. doi:10.1158/0008-5472.CAN-14-1346
- Gencik, A., Wey, W., & Muller, H. (1986). High incidence of laryngeal carcinoma in a swiss family. *ORL*, 48, 162–166.
- Gollin, S. M. (2001). Chromosomal alterations in squamous cell carcinomas of the head and neck: window to the biology of disease. *Head & Neck*, 23(March), 238–253.
- Gunduz, E., Gunduz, M., Ali, M. A. S., Beder, L., Tamamura, R., Katase, N., ... Nagatsuka, H. (2009). Loss of heterozygosity at the 9p21-24 region and identification of BRM as a candidate tumor suppressor gene in head and neck squamous cell carcinoma. *Cancer Investigation*, 27(6), 661–668. doi:10.1080/07357900802563010
- Han, J.-Y., Lee, Y.-S., Shin, E. S., Hwang, J., Nam, S., Hong, S.-H., ... Lee, J. S. (2014). A genome-wide association study of survival in small-cell lung cancer patients treated with irinotecan plus cisplatin chemotherapy. *The Pharmacogenomics Journal*, 14(1), 20–27. doi:10.1038/tpj.2013.7
- Hara, H., Ozeki, S., Shiratsuchi, Y., Tashiro, H., & Jingu, K. (1988). Familial occurrence of oral cancer: report of cases. *Journal of Oral Maxillofacial Surgery*, 46, 1098–1102. doi:10.1056/NEJM184712010371801
- Harris, L. G., Samant, R. S., & Shevde, L. a. (2011). Hedgehog signaling: networking to nurture a promalignant tumor microenvironment. *Molecular Cancer Research : MCR*, 9, 1165–1174. doi:10.1158/1541-7786.MCR-11-0175
- Hashibe, M., Brennan, P., Chuang, S.-C., Boccia, S., Castellsague, X., Chen, C., ... Boffetta, P. (2009). Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 18(February), 541–550. doi:10.1158/1055-9965.EPI-08-0347
- He, L., Lu, Y., & Guo, J. (2013). Type II VLDLR promotes cell migration by up-regulation of VEGF, MMP2 and MMP7 in breast cancer cells. *The Chinese-German Journal of Clinical Oncology*, 12(8), 374–378. doi:10.1007/s10330-013-1218-7
- He, L., Lu, Y., Wang, P., Zhang, J., Yin, C., & Qu, S. (2010). Up-regulated expression of type II very low density lipoprotein receptor correlates with cancer metastasis and has a potential link to β -catenin in different cancers. *BMC Cancer*, 10, 601. doi:10.1186/1471-2407-10-601

- Hopkins, J., Cescon, D. W., Tse, D., Bradbury, P., Xu, W., Ma, C., ... Liu, G. (2008). Genetic polymorphisms and head and neck cancer outcomes: a review. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 17(March), 490–499. doi:10.1158/1055-9965.EPI-07-2714
- Hu, L., Wu, C., Zhao, X., Heist, R., Su, L., Zhao, Y., ... Shen, H. (2012). Genome-Wide Association Study of Prognosis in Advanced Non-Small Cell Lung Cancer Patients Receiving Platinum-Based Chemotherapy. *Clinical Cancer Research*, 18, 5507–5514. doi:10.1158/1078-0432.CCR-12-1202
- Innocenti, F., Owzar, K., Cox, N. L., Evans, P., Kubo, M., Zembutsu, H., ... Ratain, M. J. (2012). A genome-wide association study of overall survival in pancreatic cancer patients treated with gemcitabine in CALGB 80303. *Clinical Cancer Research*, 18(4), 577–584. doi:10.1158/1078-0432.CCR-11-1387
- International Human Genome Sequencing Consortium. (2002). Initial sequencing and analysis of the human genome. *Nature*, 420(February), 520–562. doi:10.1038/35087627
- International HapMap Consortium. (2003). The International HapMap Project. *Nature*, 426, 789–796. doi:10.1038/nature02168
- Jemal, A., Bray, F., & Ferlay, J. (2011). Global Cancer Statistics. *CA: A Cancer Journal for Clinicians*, 61(2), 69–90. doi:10.3322/caac.20107.Available
- Kapranov, P., Willingham, A. T., & Gingeras, T. R. (2007). Genome-wide transcription and the implications for genomic organization. *Nature Reviews. Genetics*, 8(June), 413–423. doi:10.1038/nrg2083
- Laurie, C., Doheny, K., & Mirel, D. (2010). Quality control and quality assurance in genotypic data for genome-wide association studies. *Genetic Epidemiology*, 34(6), 591–602. doi:10.1002/gepi.20516.Quality
- Layland, M. K., Sessions, D. G., & Lenox, J. (2005). The influence of lymph node metastasis in the treatment of squamous cell carcinoma of the oral cavity, oropharynx, larynx, and hypopharynx: N0 versus N+. *The Laryngoscope*, 115(April), 629–639. doi:10.1097/01.mlg.0000161338.54515.b1
- Lui, V. W. Y., Hedberg, M. L., Li, H., Vangara, B. S., Pendleton, K., Zeng, Y., ... Grandis, J. R. (2013). Frequent mutation of the PI3K pathway in head and neck cancer defines predictive biomarkers. *Cancer Discovery*, 3(July), 761–769. doi:10.1158/2159-8290.CD-13-0103

- Ma, J., Tian, L., Cheng, J., Chen, Z., Xu, B., Wang, L., ... Huang, Q. (2013). Sonic Hedgehog Signaling Pathway Supports Cancer Cell Growth during Cancer Radiotherapy. *PLoS ONE*, 8(6), 1–11. doi:10.1371/journal.pone.0065032
- Marlowe, F. I. (1970). Simultaneous laryngeal tumors in sisters. *Archives of Otolaryngology*, 92, 195–197. doi:10.1001/archotol.1970.04310020093023
- Marur, S., D’Souza, G., Westra, W. H., & Forastiere, A. a. (2010a). HPV-associated head and neck cancer: a virus-related cancer epidemic. *The Lancet Oncology*, 11(8), 781–789. doi:10.1016/S1470-2045(10)70017-6
- Marur, S., D’Souza, G., Westra, W. H., & Forastiere, A. a. (2010b). HPV-associated head and neck cancer: A virus-related cancer epidemic. *The Lancet Oncology*, 11(8), 781–789. doi:10.1016/S1470-2045(10)70017-6
- Matthias, C., Branigan, K., Jahnke, V., Leder, K., Haas, J., Heighway, J., ... Hoban, P. R. (1998). Polymorphism within the cyclin D1 gene is associated with prognosis in patients with squamous cell carcinoma of the head and neck. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, 4, 2411–2418.
- McKay, J. D., Truong, T., Gaborieau, V., Chabrier, A., Chuang, S. C., Byrnes, G., ... Brennan, P. (2011). A genome-wide association study of upper aerodigestive tract cancers conducted within the INHANCE consortium. *PLoS Genetics*, 7(3). doi:10.1371/journal.pgen.1001333
- Mehanna, H., Paleri, V., West, C., & Nutting, C. (2010). Head and neck cancer — Part 1 : Epidemiology , presentation , and prevention. *BMJ*, 341, 663–666. doi:10.1136/bmj.c4684
- Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nature Reviews. Genetics*, 10(February), 155–159. doi:10.1038/nrg2521
- Mercer, T. R., & Mattick, J. S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nature Structural & Molecular Biology*, 20(3), 300–307. doi:10.1038/nsmb.2480
- Monteiro, E., Varzim, G., Pires, A. M., Teixeira, M., & Lopes, C. (2004). Cyclin D1 A870G polymorphism and amplification in laryngeal squamous cell carcinoma: Implications of tumor localization and tobacco exposure. *Cancer Detection and Prevention*, 28, 237–243. doi:10.1016/j.cdp.2004.04.005
- Mozet, C., Stoeck, M., Dimitrova, K., Dietz, A., & Wichmann, G. (2013). Hedgehog Targeting by Cyclopamine Suppresses Head and Neck Squamous Cell Carcinoma and Enhances Chemotherapeutic Effects. *Anticancer Research*, 33(6), 2415–2424.

- National Cancer Institute. (2013). National Cancer Institute SEER Cancer Statistics Review 1975-2011.
- Ohta, H., Aoyagi, K., Fukaya, M., Danjoh, I., Ohta, a., Isohata, N., ... Sasaki, H. (2009). Cross talk between hedgehog and epithelial-mesenchymal transition pathways in gastric pit cells and in diffuse-type gastric cancers. *British Journal of Cancer*, 100, 389–398. doi:10.1038/sj.bjc.6604846
- Ou, S. H. I., Zell, J. a., Ziogas, a., & Anton-Culver, H. (2007). Epidemiology of nasopharyngeal carcinoma in the United States: Improved survival of Chinese patients within the keratinizing squamous cell carcinoma histology. *Annals of Oncology*, 18, 29–35. doi:10.1093/annonc/mdl320
- Pai, S., & Westra, W. (2009). Molecular pathology of head and neck cancer: implications for diagnosis, prognosis, and treatment. *Annual Review of Pathology*, 4, 49–70. doi:10.1146/annurev.pathol.4.110807.092158.Molecular
- Pujana, M. A. (2014). Integrating germline and somatic data towards a personalized cancer medicine. *Trends in Molecular Medicine*, 20(8), 413–415. doi:10.1016/j.molmed.2014.05.004
- Pulte, D., & Brenner, H. (2010). Changes in survival in head and neck cancers in the late 20th and early 21st century: a period analysis. *The Oncologist*, 15, 994–1001. doi:10.1634/theoncologist.2009-0289
- Quintela-Fandino, M., Hitt, R., Medina, P. P., Gamarra, S., Manso, L., Cortes-Funes, H., & Sanchez-Cespedes, M. (2006). DNA-repair gene polymorphisms predict favorable clinical outcome among patients with advanced squamous cell carcinoma of the head and neck treated with cisplatin-based induction chemotherapy. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, 24(26), 4333–4339. doi:10.1200/JCO.2006.05.8768
- Radoï, L., Paget-Bailly, S., Guida, F., Cyr, D., Menvielle, G., Schmaus, A., ... Luce, D. (2013). Family history of cancer, personal history of medical conditions and risk of oral cavity cancer in France: the ICARE study. *BMC Cancer*, 13, 560. doi:10.1186/1471-2407-13-560
- Rafiq, S., Tapper, W., Collins, A., Khan, S., Politopoulos, I., Gerty, S., ... Eccles, D. (2013). Identification of inherited genetic variations influencing prognosis in early-onset breast cancer. *Cancer Research*, 73(12), 1883–1891. doi:10.1158/0008-5472.CAN-12-3377
- Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, 273(September), 1516–1517. doi: 10.1126/science.273.5281.1516

- Rubin, L. L., & de Sauvage, F. J. (2006). Targeting the Hedgehog pathway in cancer. *Nature Reviews. Drug Discovery*, 5, 1026–1033. doi:10.1038/nrd2086
- Rybicki, B. A. (2003). Allelic Loss and Tumor Pathology in Head and Neck Squamous Cell Carcinoma. *Modern Pathology*, 16, 970–979. doi:10.1097/01.MP.0000086627.74271.13
- Schlect, N. F., Franco, E. L., Pintos, J., & Kowalski, L. P. (1999). Effect of smoking cessation and tobacco type on the risk of cancers of the upper aero-digestive tract in Brazil. *Epidemiology*, 10, 412–418.
- Shu, X. O., Long, J., Lu, W., Li, C., Chen, W. Y., Delahanty, R., ... Zheng, W. (2012). Novel genetic markers of breast cancer survival identified by a genome-wide association study. *Cancer Research*, 72(5), 1182–1189. doi:10.1158/0008-5472.CAN-11-2561
- Siegel, R., Ma, J., Zou, Z., & Jemal, A. (2014). Cancer Statistics, 2014, 64(1), 9–29. doi:10.3322/caac.21208.
- Simard, E. P., Torre, L. a., & Jemal, A. (2014). International trends in head and neck cancer incidence rates: Differences by country, sex and anatomic site. *Oral Oncology*, 50, 387–403. doi:10.1016/j.oraloncology.2014.01.016
- Sims-Mourtada, J., Izzo, J. G., Apisarnthanarax, S., Wu, T. T., Malhotra, U., Luthra, R., ... Chao, K. S. C. (2006). Hedgehog: An attribute to tumor regrowth after chemoradiotherapy and a target to improve radiation response. *Clinical Cancer Research*, 12(21), 6565–6572. doi:10.1158/1078-0432.CCR-06-0176
- Stadler, Z.K., Thom, P., Robson, M.E., Weitzel, J.N., Kauff, N.D., Hurley, K.E., Devlin, V., Gold, B., Klein, R.J., Offit, K. (2010). *Journal of Clinical Oncology*, 28(27), 4255-4267. doi: 10.1200/JCO.2009.25.7816.
- Stransky, N., Egloff, A. M., Tward, A. D., Kostic, A. D., Cibulskis, K., Sivachenko, A., ... Grandis, J. R. (2011). The mutational landscape of head and neck squamous cell carcinoma. *Science*, 333(6046), 1157–1160. doi:10.1126/science.1208130
- Streit, S., Bange, J., Fichtner, A., Ihrler, S., Issing, W., & Ullrich, A. (2004). Involvement of the FGFR4 Arg388 allele in head and neck squamous cell carcinoma. *International Journal of Cancer*, 111, 213–217. doi:10.1002/ijc.20204
- Suh, Y., Amelio, I., Guerrero Urbano, T., & Tavassoli, M. (2014). Clinical update on cancer: molecular oncology of head and neck cancer. *Cell Death & Disease*, 5(1), e1018. doi:10.1038/cddis.2013.548

- Tang, S., Pan, Y., Wang, Y., Hu, L., Cao, S., Chu, M., ... Shen, H. (2015). Genome-wide Association Study of Survival in Early-stage Non-Small Cell Lung Cancer. *Annals of Surgical Oncology*, 22, 630–635. doi:10.1245/s10434-014-3983-0
- The Cancer Genome Atlas Network. (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*, 519, 576–582. doi:10.1038/nature14129
- Trizna, Z., & Schantz, S. P. (1992). Hereditary and environmental factors associated with risk and progression of head and neck cancer. *Otolaryngologic Clinics of North America*, 25, 1089–1103.
- Villegas, V., & Zaphiropoulos, P. (2015). Neighboring Gene Regulation by Antisense Long Non-Coding RNAs. *International Journal of Molecular Sciences*, 16, 3251–3266. doi:10.3390/ijms16023251
- Wang, J. R., Gramling, S. J. B., Goldstein, D. P., Cheng, D., Chen, D., Azad, A. K., ... Liu, G. (2013). Association of two BRM promoter polymorphisms with head and neck squamous cell carcinoma risk. *Carcinogenesis*, 34(5), 1012–1017. doi:10.1093/carcin/bgt008
- Wong, K., Huang, S. H., O'Sullivan, B., Lockwood, G., Dale, D., Michaelson, T., ... Ringash, J. (2010). Point-of-care outcome assessment in the cancer clinic: audit of data quality. *Radiotherapy and Oncology : Journal of the European Society for Therapeutic Radiology and Oncology*, 95(3), 339–343. doi:10.1016/j.radonc.2010.03.015
- Wu, C., Kraft, P., Stolzenberg-Solomon, R., Stepilowski, E., Brotzman, M., Xu, M., ... Wolpin, B. M. (2014). Genome-wide association study of survival in patients with pancreatic adenocarcinoma. *Gut*, 63, 152–160. doi:10.1136/gutjnl-2012-303477
- Wu, X., Wang, L., Ye, Y., Aakre, J. a., Pu, X., Chang, G. C., ... Yang, P. (2013). Genome-wide association study of genetic predictors of overall survival for non-small cell lung cancer in never smokers. *Cancer Research*, 73, 4028–4038. doi:10.1158/0008-5472.CAN-12-4033
- Wyss, A. B., Weissler, M. C., Avery, C. L., Herring, A. H., Bensen, J. T., Barnholtz-Sloan, J. S., ... Olshan, A. F. (2014). Single nucleotide polymorphisms in nucleotide excision repair genes, cancer treatment, and head and neck cancer survival. *Cancer Causes & Control : CCC*, 25, 437–450. doi:10.1007/s10552-014-0346-x
- Xu, L., Doan, P. C., Wei, Q., Liu, Y., Li, G., & Sturgis, E. M. (2012). Association of BRCA1 functional single nucleotide polymorphisms with risk of differentiated thyroid carcinoma. *Thyroid : Official Journal of the American Thyroid Association*, 22(1), 35–43. doi:10.1089/thy.2011.0117

- Xu, W., Xu, J., Shestopaloff, K., Dicks, E., Green, J., Parfrey, P., ... Savas, S. (2015). A genome wide association study on Newfoundland colorectal cancer patients' survival outcomes. *Biomarker Research*, 3, 6. doi:10.1186/s40364-015-0031-6
- Yang, J. J., Cheng, C., Devidas, M., Cao, X., Campana, D., Yang, W., ... Relling, M. V. (2012). Genome-wide association study identifies germline polymorphisms associated with relapse of childhood acute lymphoblastic leukemia. *Blood*, 120(20), 4197–4204. doi:10.1182/blood-2012-07-440107.

Supplementary Information

Figure 1. Study overview

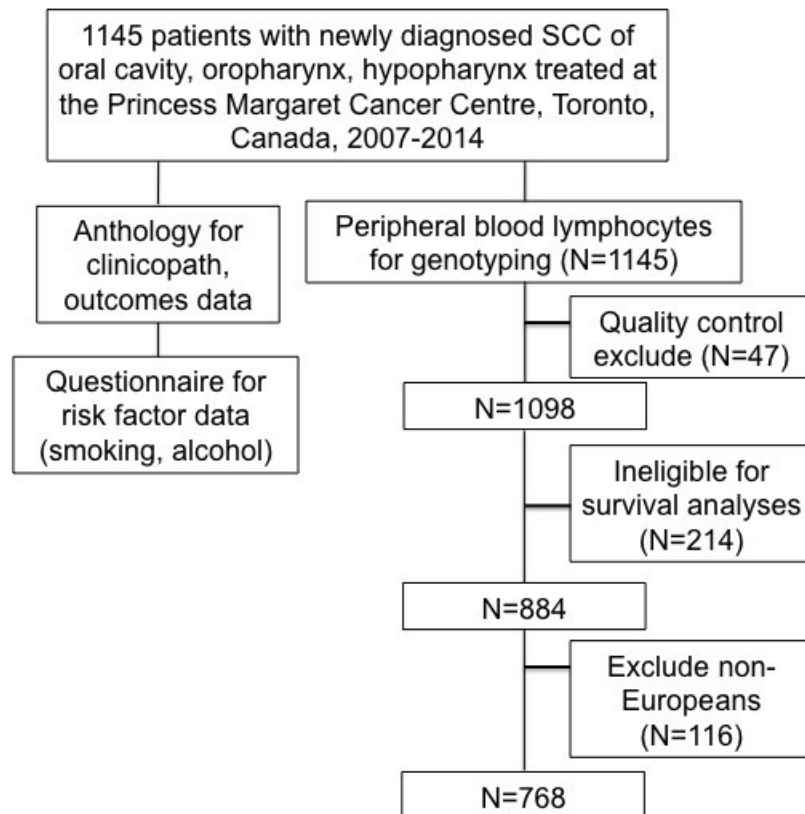


Figure 2. Pairwise IBD before and after removal of related pairs

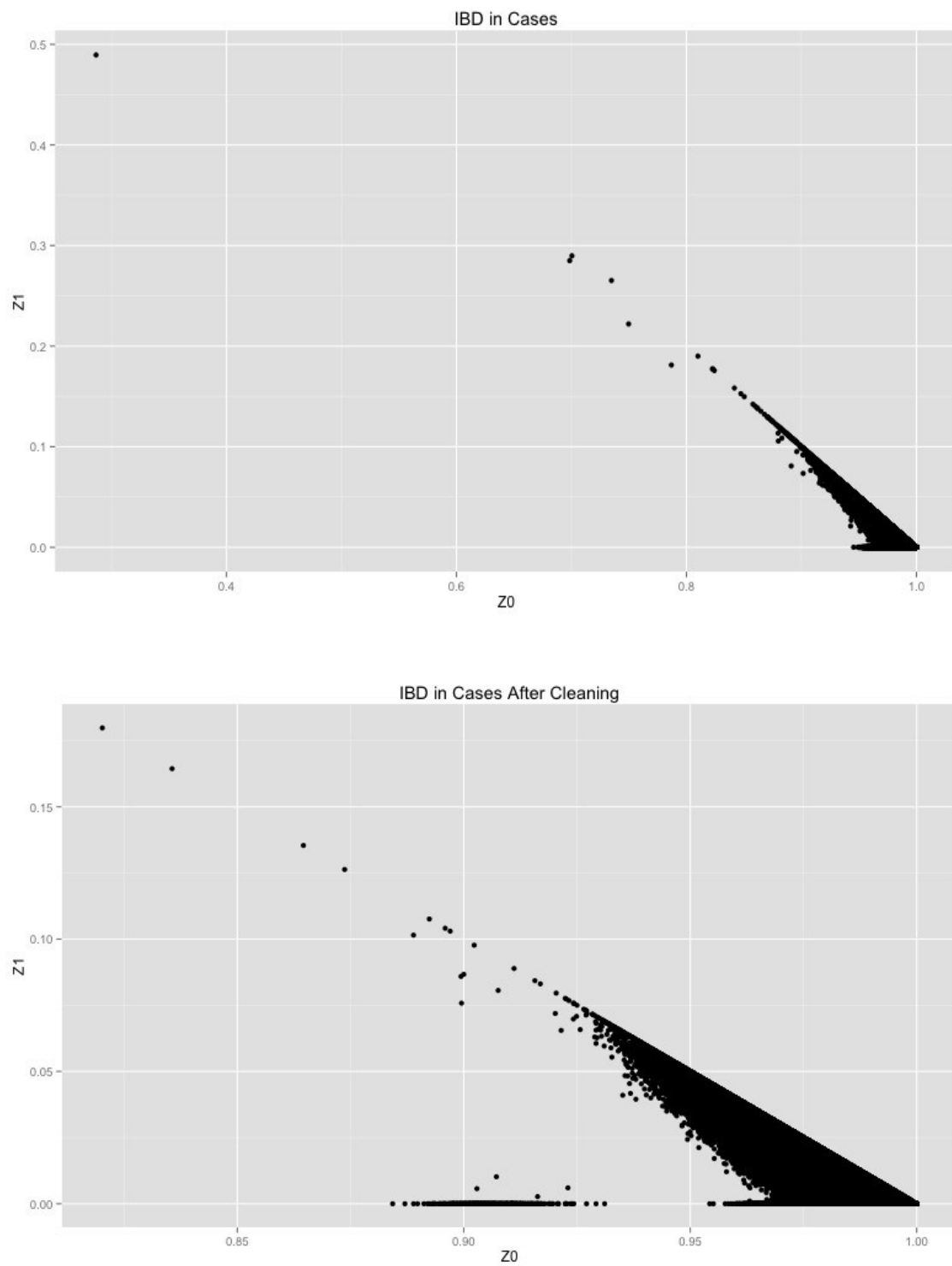


Figure 3. Distribution of sample call rate before and after sample-based QC

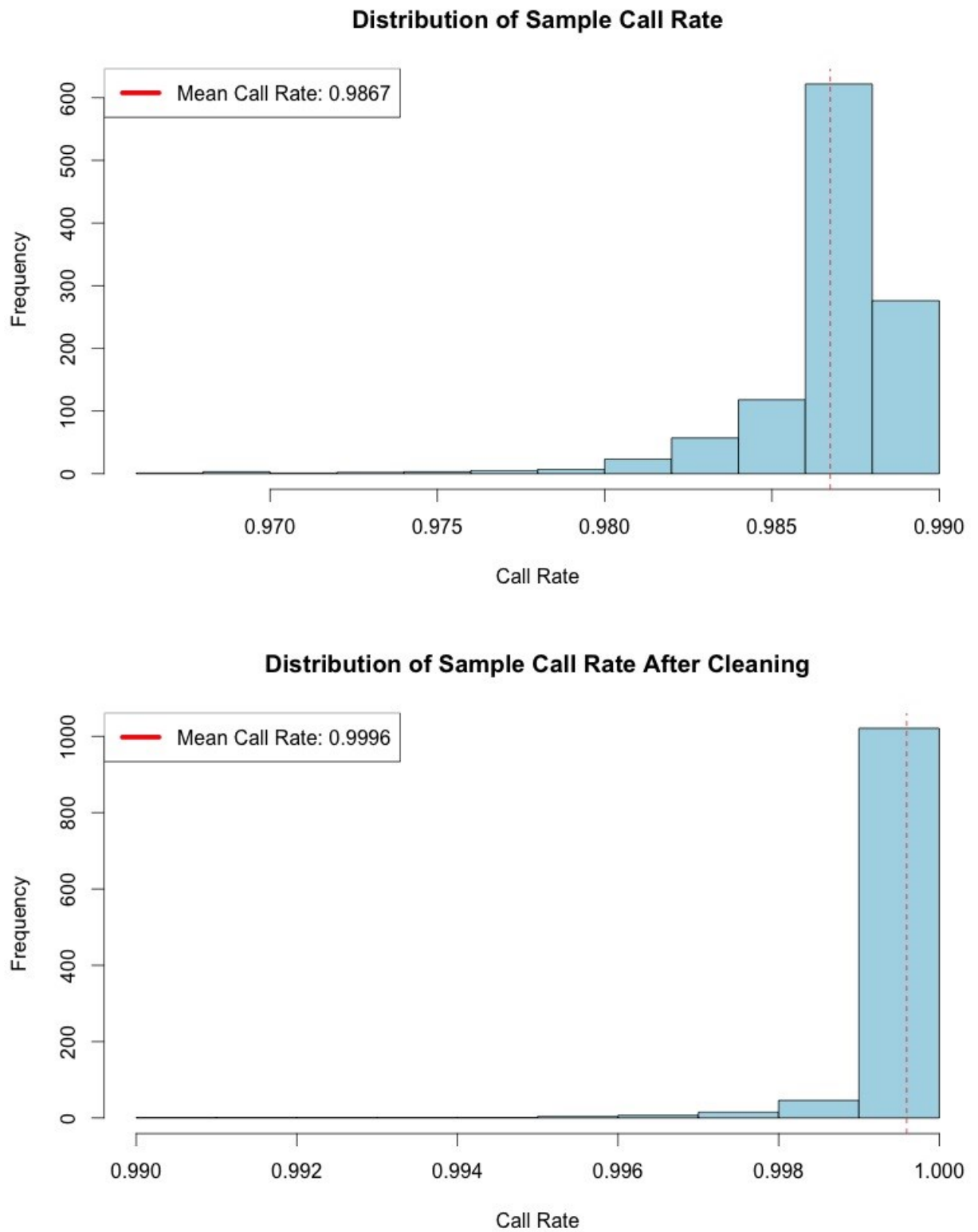


Figure 4. Distribution of SNP call rate before and after marker-based QC

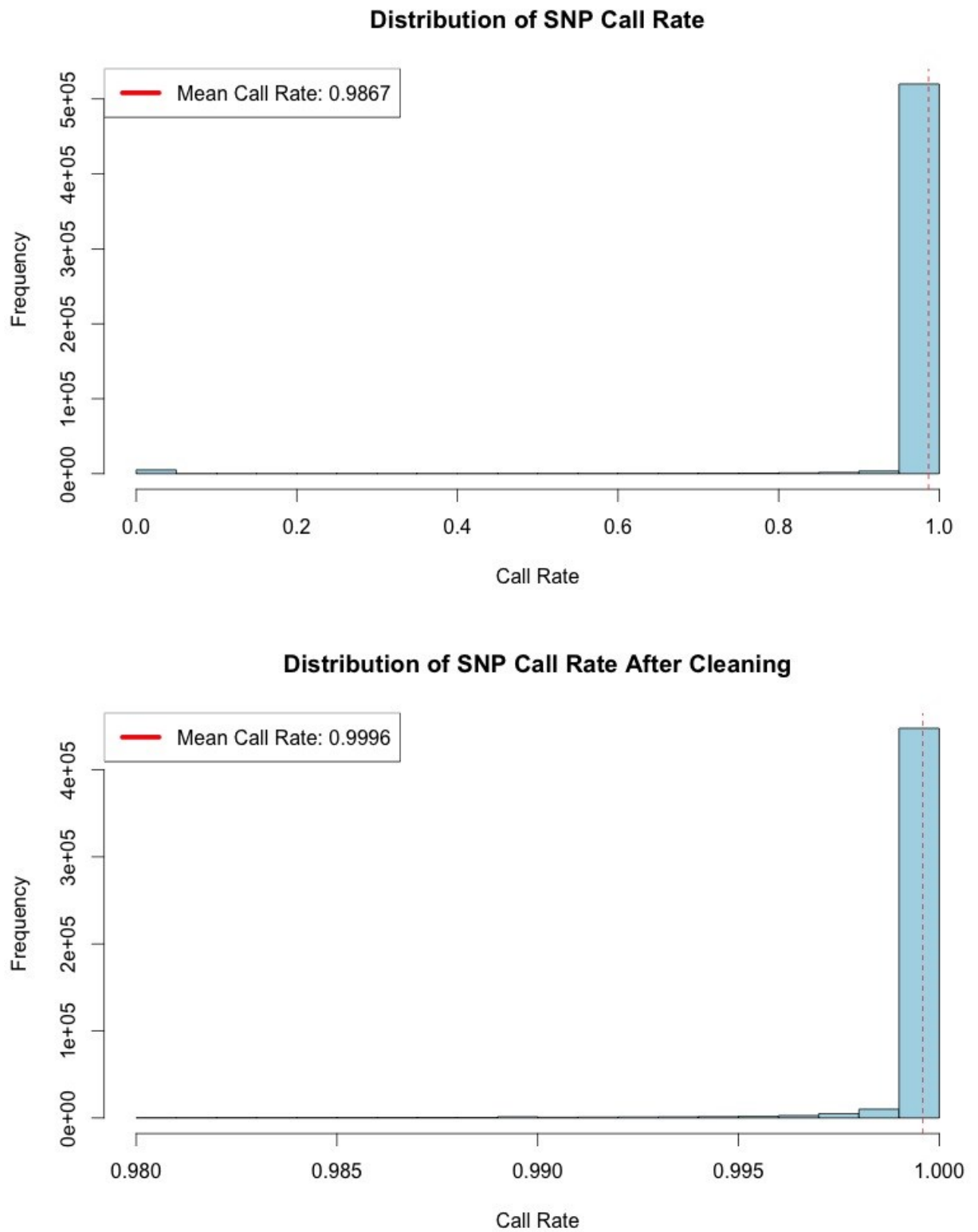
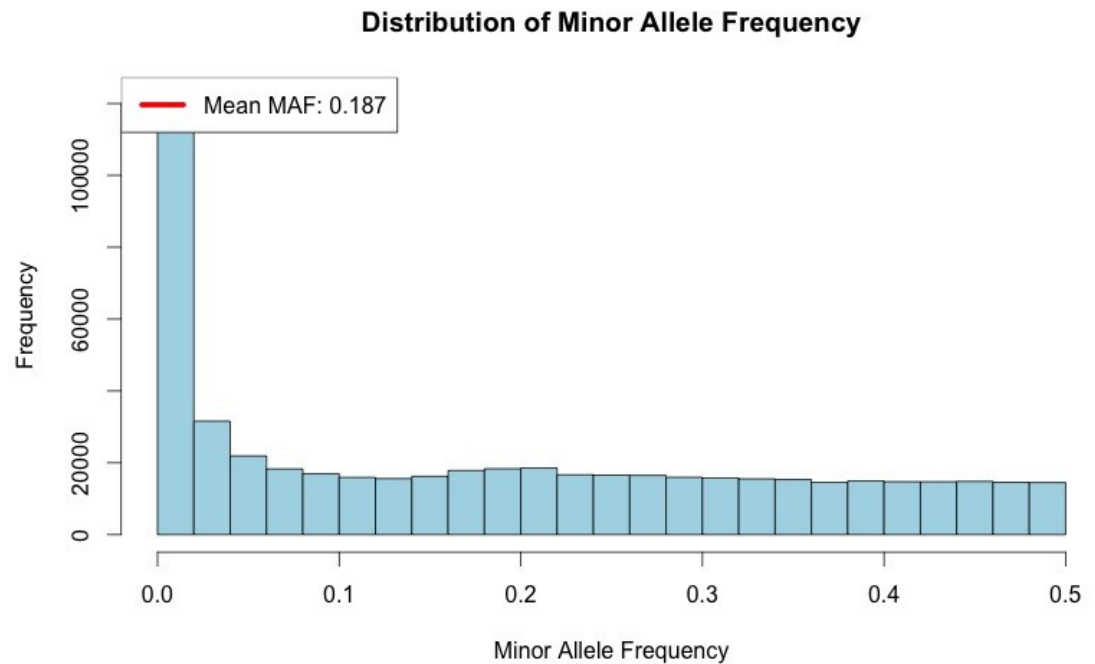


Figure 5. Distribution of minor allele frequencies before and after SNP and sample-based QC

a) Before QC



b) After QC

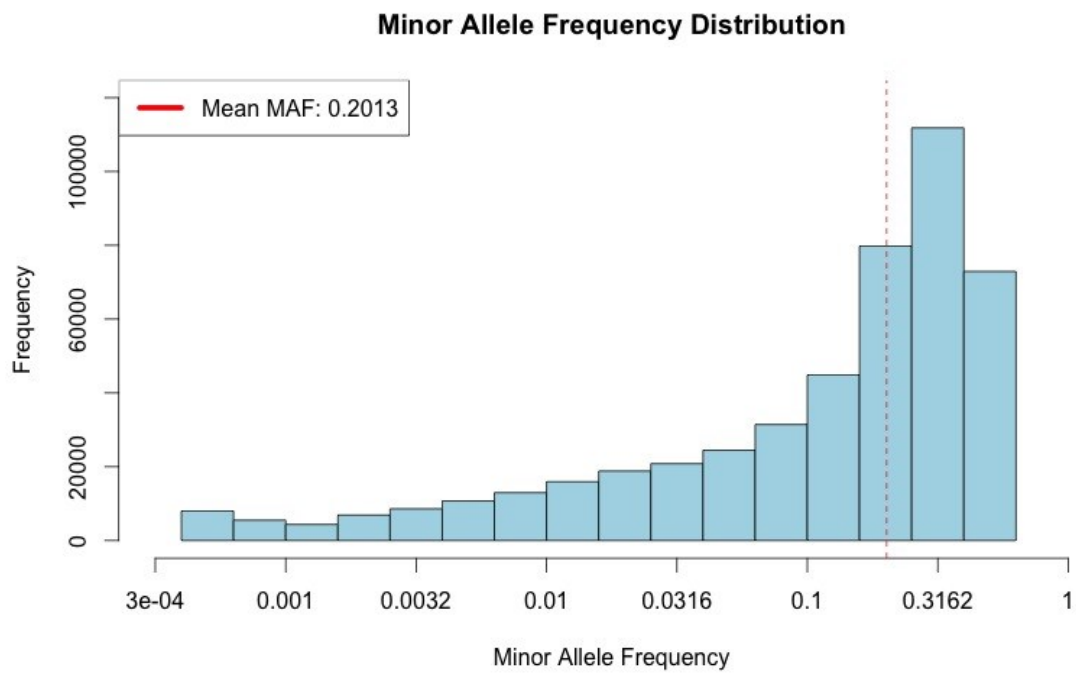
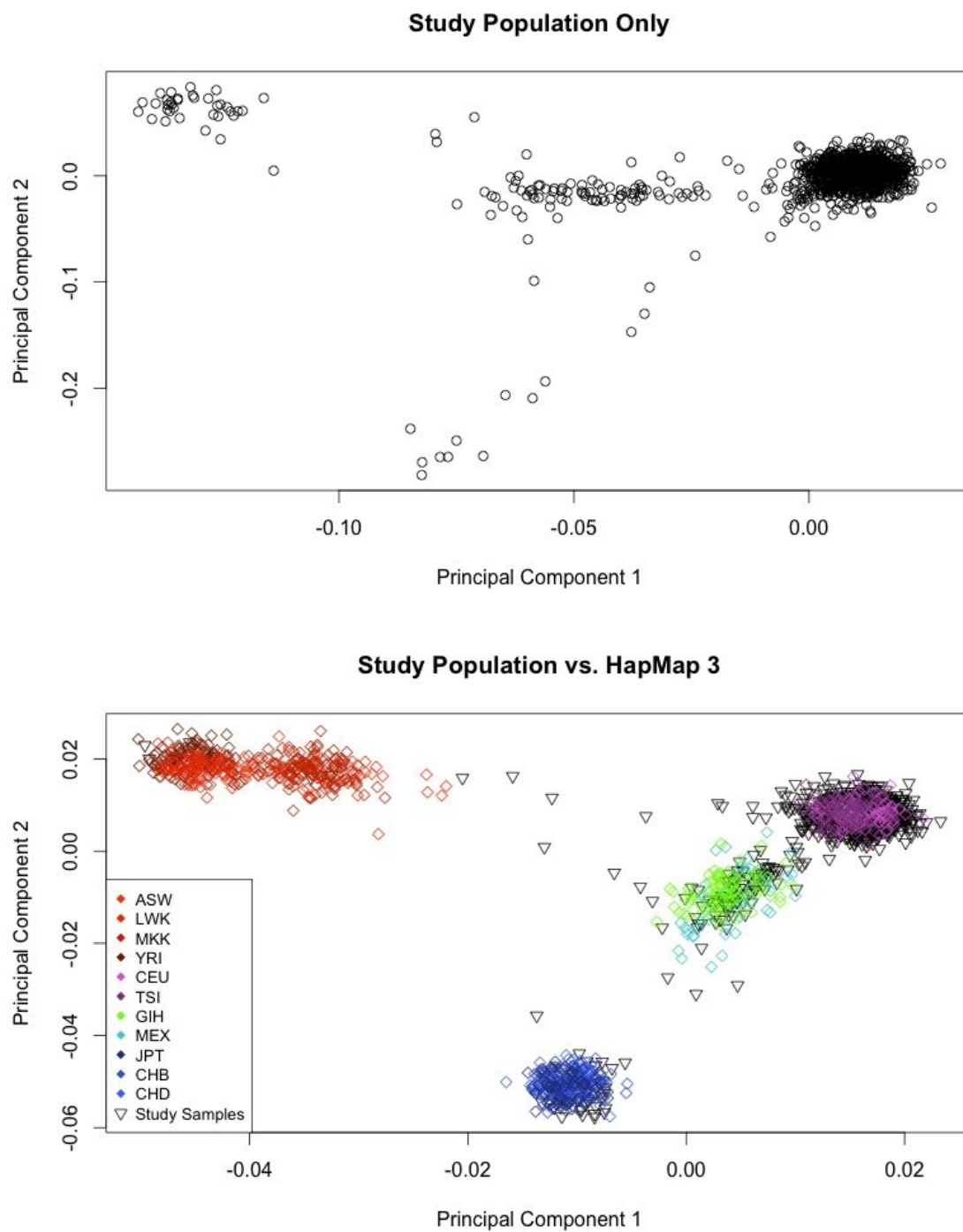
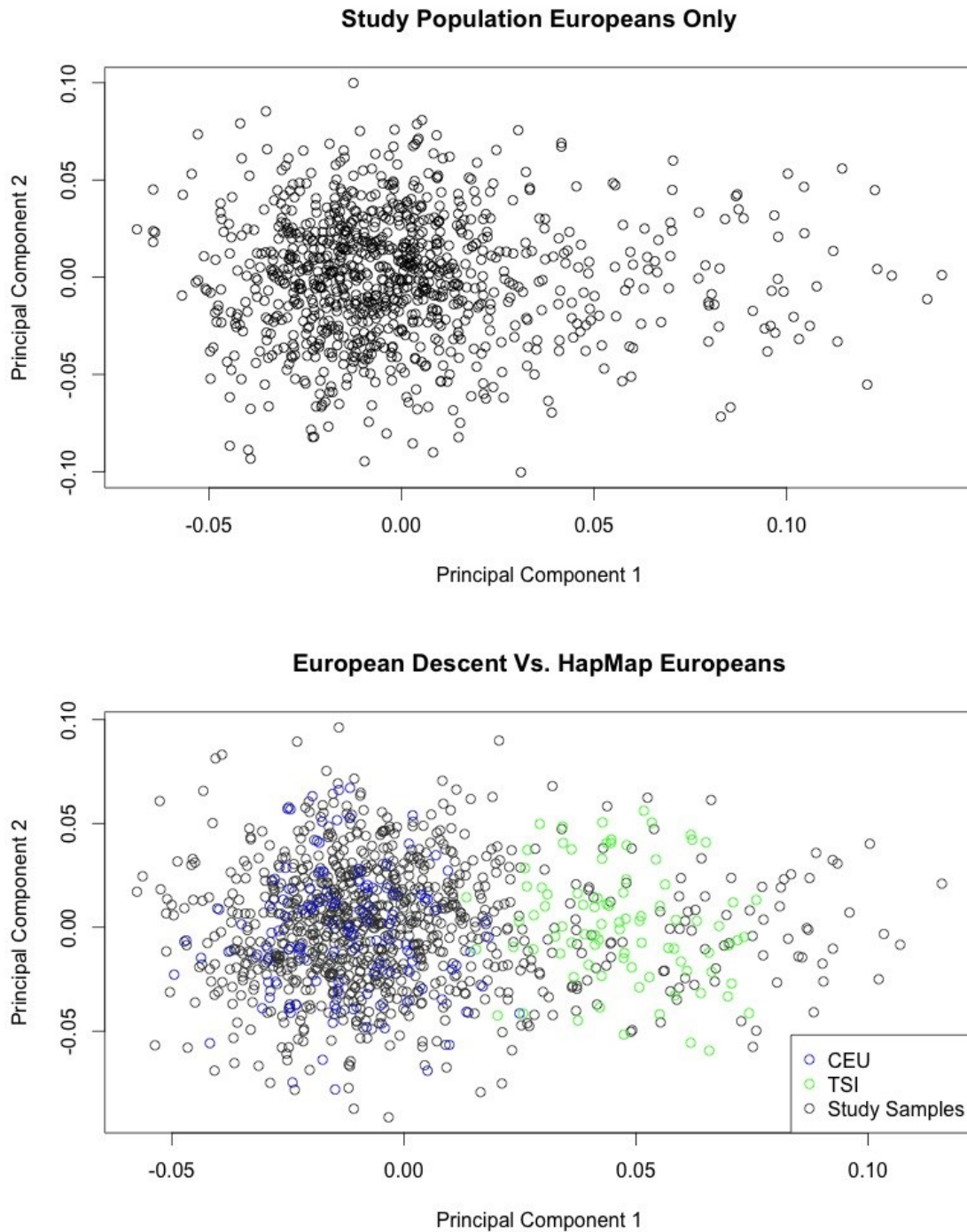


Figure 6. First two principal components of the study population with and without reference HapMap 3 populations



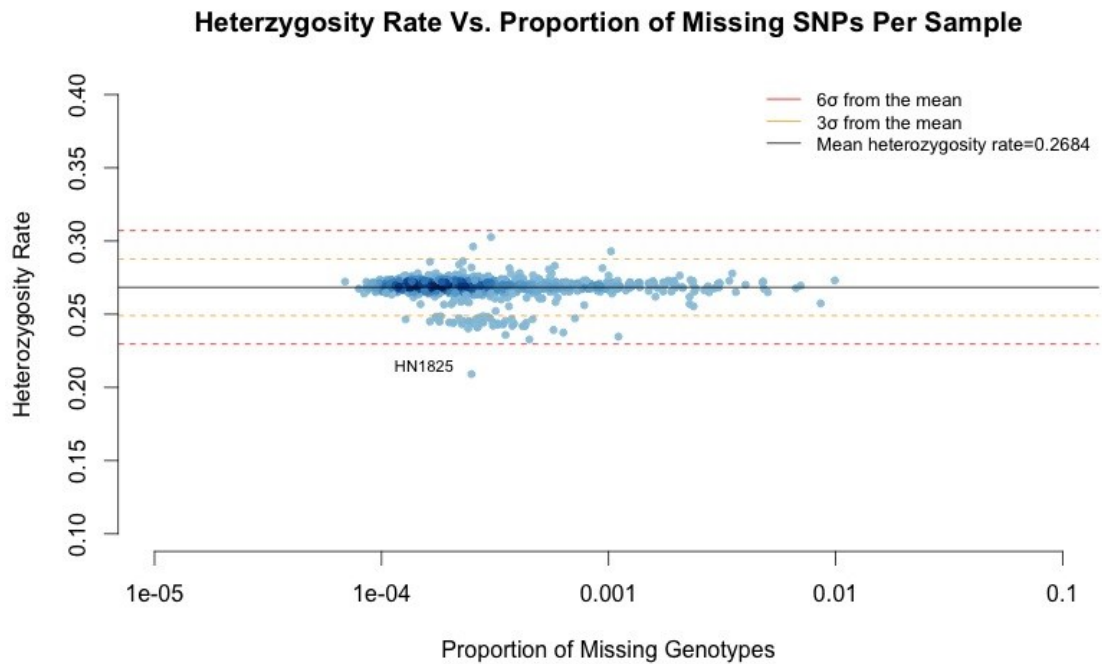
*1837 Ancestral Informative Markers (AIMS) included on the custom OncoArray were used to generate principal components. Population outliers were excluded.

Figure 7. First two principal components of the study population Europeans with and without reference HapMap 3 CEU and TSI populations



*1837 Ancestral Informative Markers (AIMS) included on the custom OncoArray were used to generate principal components.

Figure 8: Heterozygosity rate vs. proportion of missing SNPs per sample after SNP and sample-based QC



*one sample was excluded based on heterozygosity rate greater than 6 standard deviations from the mean

Figure 9. Minor allele frequencies and sample call rates across genotyping plates

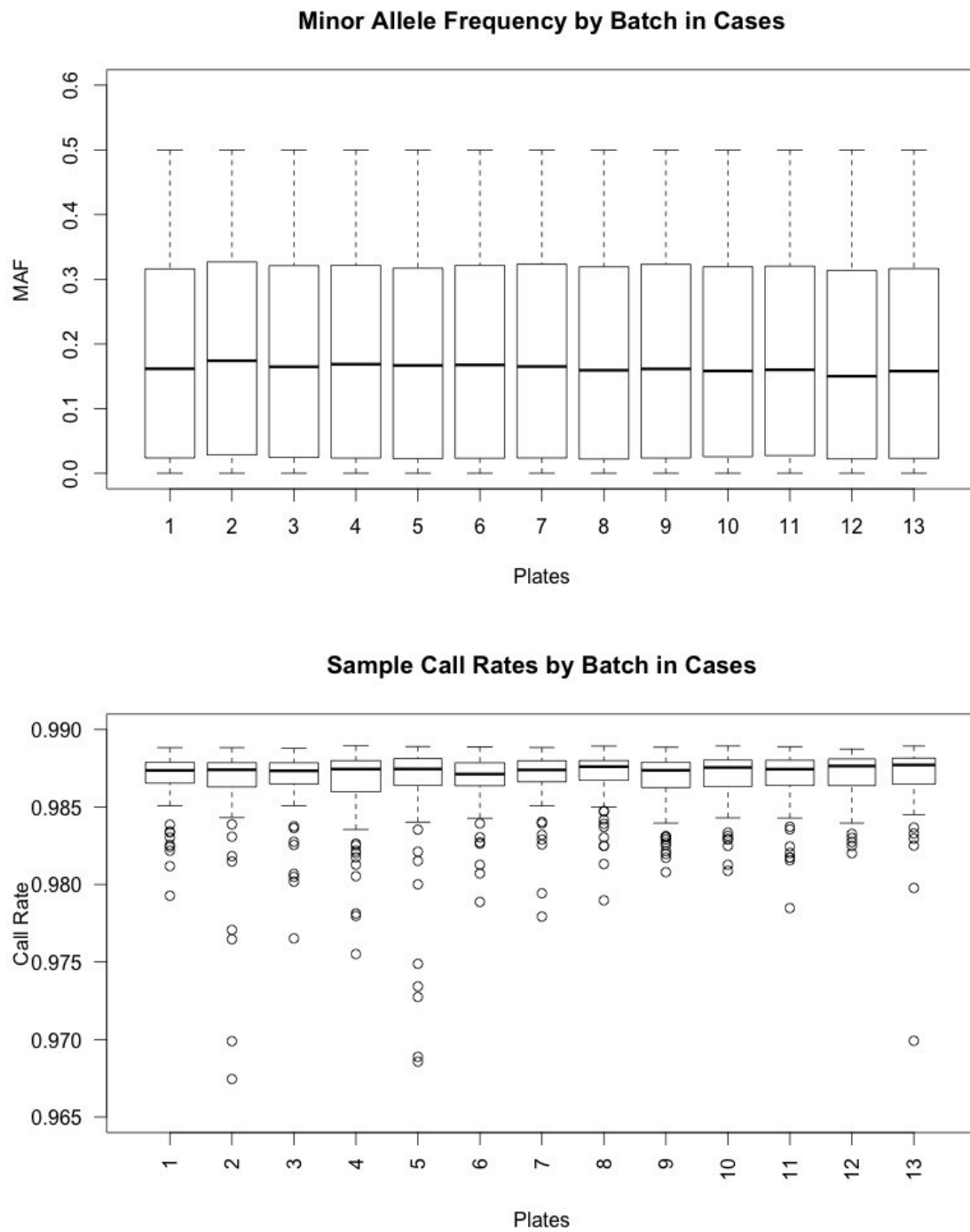
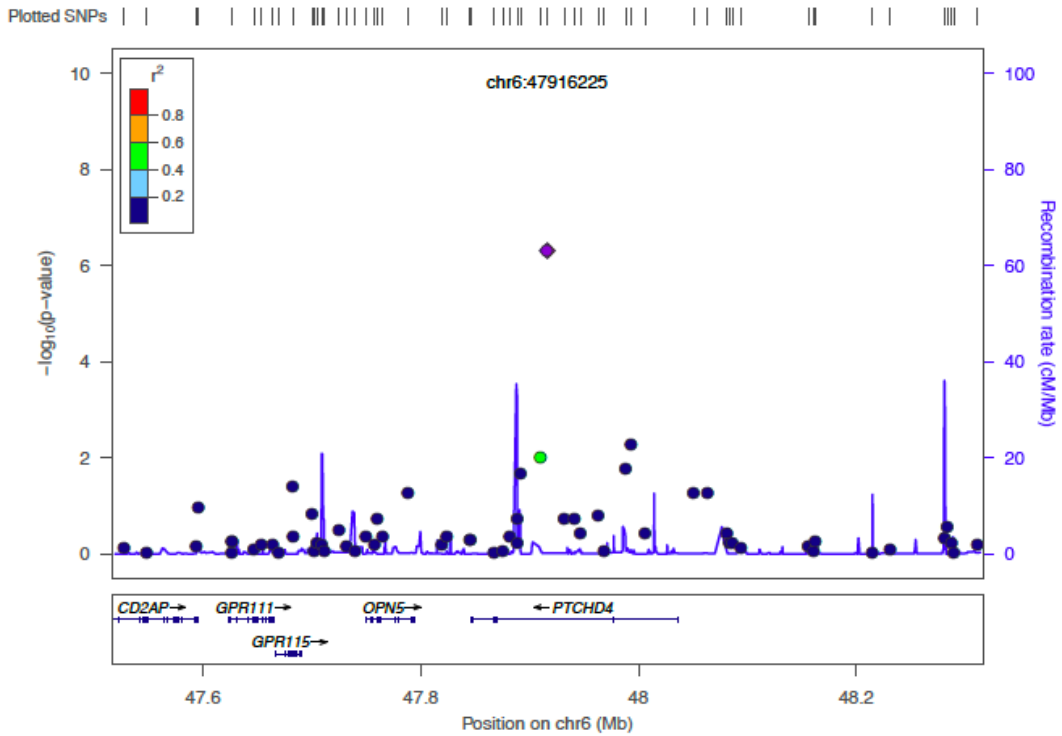
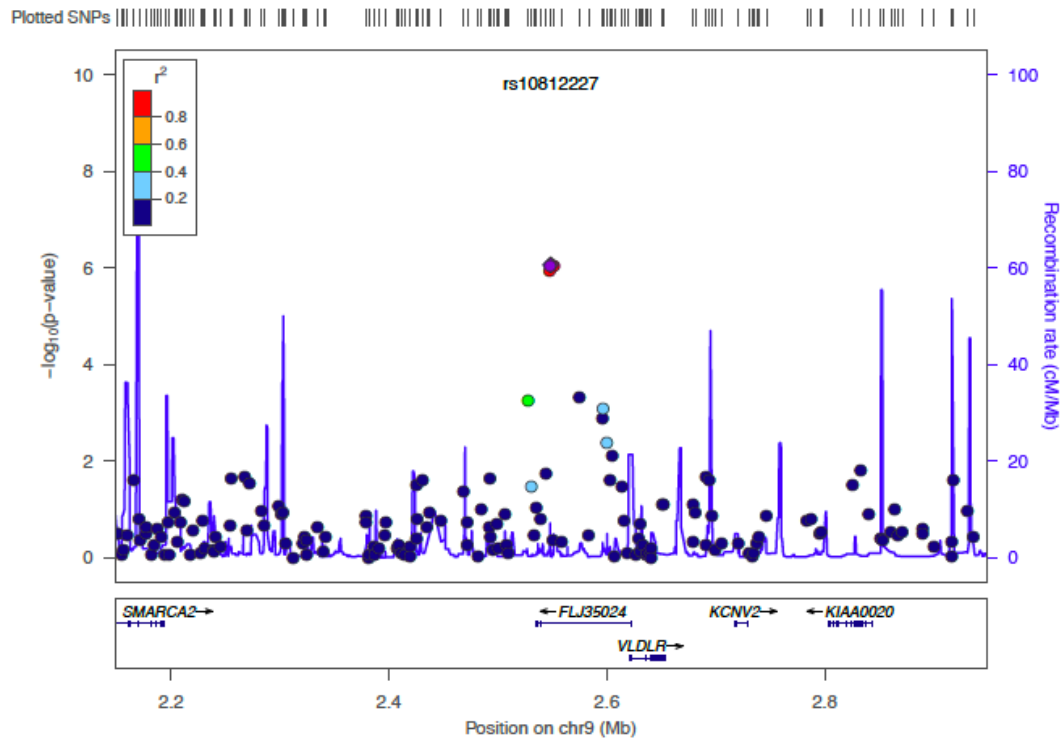


Figure 10. Regional plots for top snps across all anatomical sub-sites
a) *rs1974051*



b) *rs10812227* and *rs12237653*



**rs10812227* and *rs12237653* are in high LD with $r^2 > 0.8$. *FLJ35024* is also known as *VLDLR-AS1*.

Figure 11. Regional plot for the most significant SNP in oropharynx sub-site

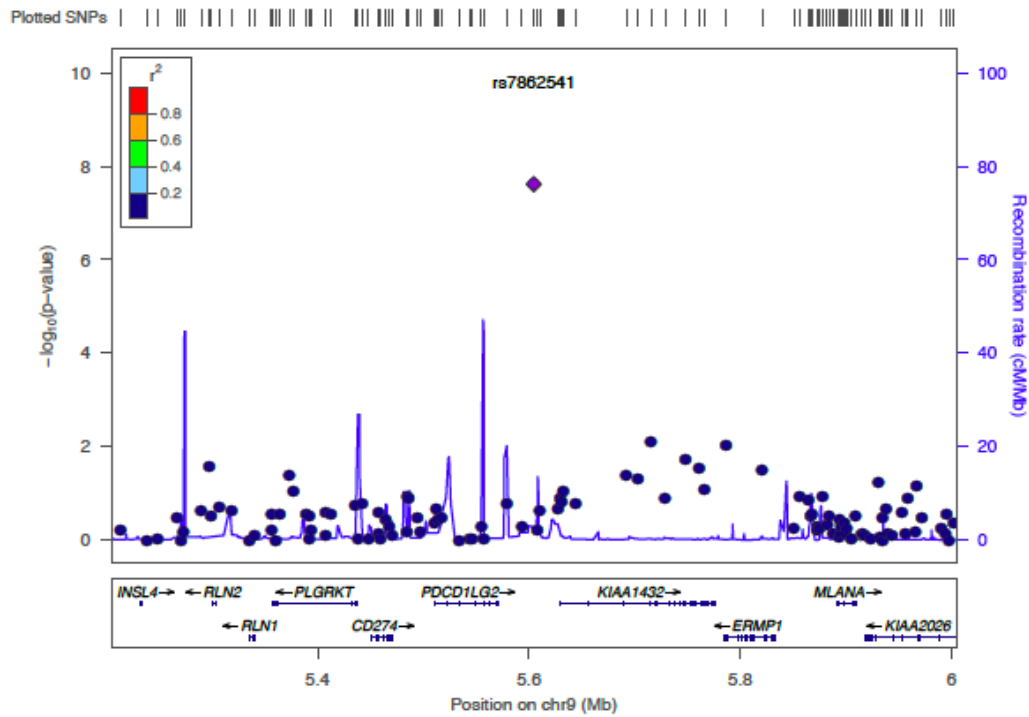
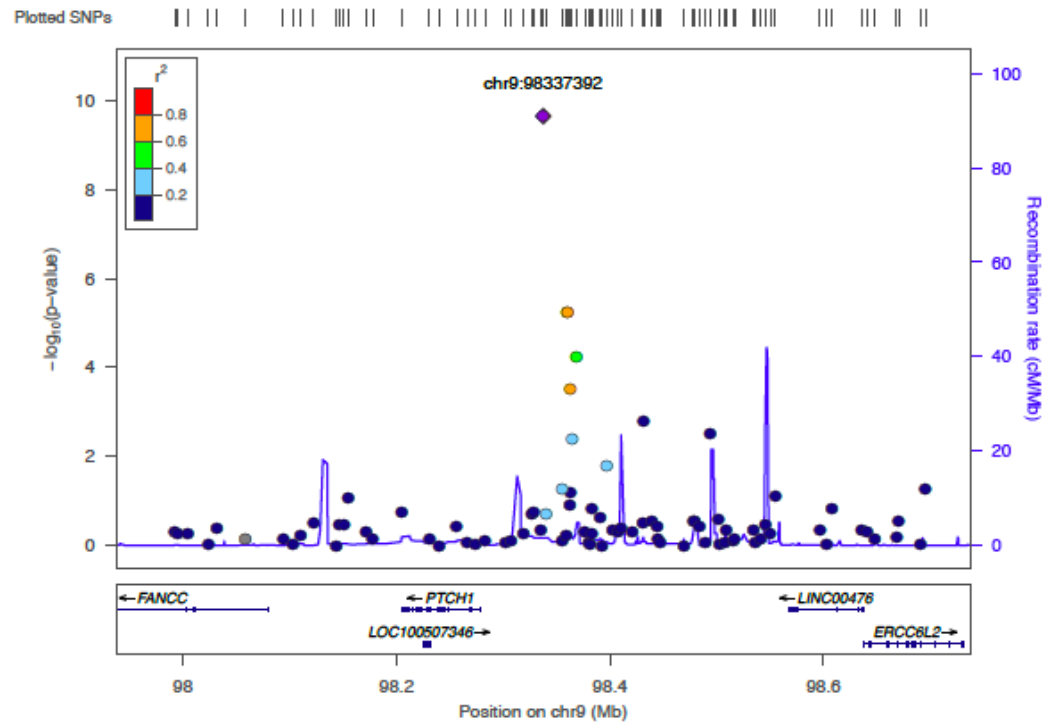


Figure 12. Regional plot for the most significant SNP in the oral cavity sub-site



Curriculum Vitae

NAME OF AUTHOR: Jennifer Rui Wang

DATE AND PLACE OF BIRTH: July 27, 1984, Guiyang, China

DEGREES AWARDED:

Bachelor of Health Sciences (Honours), McMaster University, Hamilton, Canada, 2006

Doctor of Medicine, University of Toronto, Toronto, Canada, 2010

PROFESSIONAL EXPERIENCE:

Postgraduate Residency Training in Otolaryngology-Head and Neck Surgery, University of Toronto, Toronto, Canada, 2010-2017

Clinician-Investigator Program, University of Toronto, Toronto, Canada, 2013-2015

HONORS AND AWARDS:

Ballantyne Resident Research Grant, American Head and Neck Society, 2014-2015

Chapnik, Freeman & Friedberg Clinical Scientist Program Award, Department of Otolaryngology-Head and Neck Surgery, University of Toronto, 2014-2015

William S. Fenwick Research Fellowship, University of Toronto, 2014-2015

Harry Barberian Research Scholarship, Department of Otolaryngology-Head and Neck Surgery, University of Toronto, 2013-2015

Joseph M. West Family Memorial Fund, University of Toronto, 2014

Master's Tuition Scholarship, Johns Hopkins Bloomberg School of Public Health, 2014

Hugh Barber Scholarship in Otolaryngology, Department of Otolaryngology-Head and Neck Surgery, University of Toronto, 2010